

Reforming Stalin's academies

One of the most obvious relics of Stalinism is the continued survival in Eastern Europe of academies of science whose purposes have never been fulfilled. This is a time for changing them.

LENIN, to his credit, was a great believer in the social and economic value of research, as a result of which the Soviet Union is spattered with academies of science. There is one for each of the 16 autonomous republics as well as the Academy of Science of the USSR, the 'big academy', transplanted reluctantly from Leningrad to Moscow in the 1930s. The idea was that the academies would take charge of most of what counts as serious research, while separate parts of the bureaucracy would be responsible for the education of students and the development of industry. With the passage of time, it has become plain that the arrangement is defective, and should be changed — quite how that will be done in the Soviet Union remains to be seen. By contrast, recent events in Eastern Europe create opportunities for change that should be quickly seized.

Stalin, hardly the most reflective of men, seems to have given little thought to the decision, after 1945, that Soviet-style academies should be replicated throughout Eastern Europe, which remains the general pattern (but Romania managed to keep social sciences separate). In Czechoslovakia and now in Poland (see page 108), the new governments of the past few weeks have changed the membership of their academies' praesidia (governing councils), but the institutions themselves remain for the time being. But their influence on national affairs has, mercifully, not always been as powerful as that of the Soviet Union. In Poland, for example, universities have retained a substantial measure of independence in research, while several Hungarian research institutes have kept intact their own links with the outside world through difficult times. Yet now, even the hard-pressed and often apparently provisional governments of Eastern Europe should jump at the chance for change with which they are presented.

The objection to the standard Soviet model of an academy, amply borne out by the experience of the past 40 years, is that it is self-contradictory. People are appointed members of the academy on the strength of their distinction as researchers and scholars, but are then required collectively to manage a substantial part of their government's investment in science and technology — a task for which, individually, they may have no talent. So, inevitably, governments, taking Lenin's view that science and technology are important, have concluded that they are also too important to be left to the possibly wayward

care of those whom their academies elect to run their own affairs. The result is that even where, in Eastern Europe, academics individually have been able to hang onto a measure of independence, their academies have necessarily been tools of government.

That is what most of all needs changing, and there is a simple way in which it should be done. Especially in Eastern Europe, governments need both the advice an independent academy can offer (would Poland's environmental problems have been as serious now if its academy had been more free ten years ago?) and a means of channelling energy into applied research. That argues for a simple separation of powers between the model academy (which could even be asked, for the time being, to administer research grants to universities) and the execution of industrial research policy, which should be the responsibility of a government department for as long as public funds are spent. On the face of things, such changes will seem a retreat from utilitarian values in research, but in reality they are an opportunity for more pointed research.

A more challenging goal would be to end the present isolation from each other of national research enterprises. Certainly it makes no sense that each of them should support separate agricultural and medical research establishments (run by separate national academies), but there are endless opportunities for collaboration in other fields. If they move quickly enough in these directions, the new governments might even see the academies to their East follow suit, which would be a piquant triumph. □

Adirondacks awake

A new wave of cataclysmic writing, now about the greenhouse effect, threatens to engulf us.

EVERY now and again, the *New Yorker*, in its devotion to fine writing, plays a cruel trick on its readers by commissioning a fine writer (in every sense of the word 'fine') to give an account of the latest fashion in technological cataclysm. That is the route by which, in 1963, Rachel Carson's *Silent Spring* came to startle the world and how Jonathan Schell's vision of nuclear Armageddon kept nervous New Yorkers (if there are such) awake at nights four years ago. Now it is the turn of the greenhouse effect. Bill McKibben's series of articles on this theme last



year has now been published in book form as *The End of Nature* (Viking-Penguin, £12.99 in Britain) and will be more formally reviewed in a forthcoming issue. The immediate question is whether McKibben's view of nature, a kind of extrapolation of Thoreau's Walden Pond, will help or hinder rational strategies for fending off climatic change.

McKibben has lived for two years in the Adirondacks, in upper New York State. His readers are fully informed of the delectable details of that rural life — how long it takes to cycle to the post office six miles away and so on. Plainly, he has found an enviable spot:

A forty-minute hike brings the dog and I to the top of the hill behind my home. There's a ledge there that offers a fine view in winter when the leaves are down. I know the hill well by now, each gully and small creek . . . I know the places where the deer come, and the coyotes after them . . . I can see my house down below, white against the hemlocks. I can see my whole material life — my car, the bedroom, the chimney above the stove. But a choice seems unavoidable. Either that life down there changes, perhaps dramatically, or this life all around me up here changes — passes away.

But McKibben is not an ordinary ecodoomsayer. He castigates earlier prophets of that tradition, Paul Ehrlich for example, for having spoiled their arguments by exaggeration. He accepts that hydroelectric schemes would take the edge off the greenhouse effect, but weeps that they will sink the knife deeper into nature. He worries that he and his wife may have to live closer to other people for the sake of energy-conservation. His view is that he and the rest of us have been cheated — one chapter is headed "A promise broken". But of what? An uncomplicated life, it seems.

Whenever, McKibben should reflect, was it otherwise? Even in Thoreau's time, the contemplation of one's house, whether "white against the hemlocks" or some other colour, was a rare luxury for most people. Now, especially in a United States beset by social problems as wild as the appetite for narcotics and the poverty of public education, it is a kind of self-indulgence. None of that implies that the disappearance of the hemlocks would be welcome, or that global warming occasioned by the greenhouse effect would not be a serious matter compelling compromises of the way we now live. But it serves no better purpose than the exaggerations of the 1970s to pretend that the greenhouse effect is the first global threat, to yearn for the avoidance of compromise and to fail to say what is being done. □

Protecting chips

Europe hankers after protection for chip manufacturers, thereby damaging Europe's economy and its reputation.

MOUSETRAP manufacturers the world over are not at present seriously threatened by imports of their products from Japan but, if (and when?) they are, they would probably not respond by demanding that Japanese mousetrap manufacturers should be made even more profitable than at present. If mousetrap manufacturing

were a business in which new capital investment mattered, it would certainly be rash to arrange that competitors' profit margins should be artificially increased. The clever Japanese would then be even better placed to pour resources into the design of better mousetraps and also to recruit more robots for their mousetrap production lines. Mousetrap manufacturers would have only two ways of responding to the threat from Japan — to become more efficient or to go out of business. That is how it should be.

But that is not how it is with electronic chips, the integrated circuit elements that have become the essential ingredients of computer construction. Two years ago, the United States reached a controversial agreement with Japanese manufacturers of memory chips that sets minimum prices for the sale of Japanese-made products in the United States. Now Europe appears bent on following suit. The European Commission has apparently reached an agreement with 11 Japanese manufacturers that their memory chips should not be sold at prices less than certain minima. The objective is to give European manufacturers of the same products a protected breathing space. More complicated chips, microprocessors for example, are not affected.

There are several things to be said about this plan, of which the chief is that it will not achieve its intended purpose. Everybody knows that chip-manufacturing requires huge investments in micromanipulative technology of various kinds and that success goes to those who can operate the machines they buy so as to keep the proportion of unusable products to a minimum. The first task takes cash, the second experience. Most European manufacturers, taken by surprise by the explosion of this technology in the past 15 years, are trying to catch up by seeking niche markets in which specialized chips can command prices high enough to justify the capital investment as well as to accommodate their own learning curves. Most manufacturers, even the Japanese, envy companies such as the US Intel Corporation that command the mass market for integrated microprocessor chips.

The well-known danger in the European plan is that the negotiated breathing space will allow European manufacturers to pretend to compete in the toughest part of the microchip market, encouraging them to clamour for more protection when the arrangement ends (in 1991, as now planned). By then, Japanese manufacturers, whose market-share is unlikely to be much affected in the near future, will be able to show European manufacturers an even cleaner pair of heels. Meanwhile, the costs of the new arrangement will be met by others — those in Europe who manufacture computers and their customers.

The analogy with mousetrap manufacture is in this sense inexact. A mousetrap is an end in itself, but a memory chip in a computer is a means to a larger end. By planning to protect chip manufacturers from the Japanese, the European Commission is implicitly choosing between the economic value of chip-manufacturing and that of the availability of cheap computers. It is choosing wrongly. □

Price tag for accelerator jumps by \$1,000 million

■ Panel of physicists called in to provide support

■ Machine at old price not worth building

Washington

AFTER months of rumours, US Department of Energy (DoE) officials have admitted that the Superconducting Super Collider (SSC), the 53-mile-circumference, 20 TeV proton-proton collider, will cost a lot more than previously advertised. But Secretary of Energy Admiral James

that an SSC of 20 TeV will be able to find it. What the DoE panel must decide is how far the energy can be cut without losing the machine's *raison d'être*.

Some panel members have already made their opinions known. "If you miss the Higgs because you went to 17 TeV, that's a tragedy", says Nobel laureate and panel member Leon Lederman.

Fermilab Schwitters also takes a hard line: "Anything below a 10 per cent cut in energy and you get a situation that would be a disaster. And anything better than a 10 per cent cut costs more than \$5.9 billion." Only last month, Watkins said that the \$5,900 million figure was "fixed". "Unless I am told by . . . solid peer review . . . why it needs to be more than that, I don't see any reason to talk about cost increases at this time", he said.

The sudden change not only promises to enliven the SSC's numerous Congressional critics, especially with President George Bush's 1991 budget due out on 29 January, but also raises the ghost of Brookhaven National Laboratory's 'Isabelle' accelerator, which was cancelled in 1983 after construction had begun. By asking the review panel to submit its findings two weeks before the 1991 budget is presented, Watkins' aim is to be ready with a strong scientific message to counter the bad news on the SSC's price, Schwitters says.

The SSC's technical troubles come mostly from its superconducting magnets. The diameter ('aperture') of the beam pipe through which the protons will travel, and the energy of the injector, a smaller accelerator which feeds protons into the main ring, were both predicated on the main ring magnets achieving a magnetic field of sufficiently high uniformity. But prototype SSC magnets have not been performing as well as expected, and to compensate for these shortcomings SSC designers now want to increase the aperture from 4 cm to 5 cm (so that the beam will not hit the walls), and to raise the injector energy from 1 TeV to 2 TeV.

The increase in aperture means bigger magnets, and raises the price by \$300 million. The bigger injector adds another \$288 million. And whereas a 1-TeV injector could be modelled fairly directly on the 1 TeV Tevatron at the Fermi National

Accelerator Laboratory, the world's first and only superconducting proton accelerator, a 2-TeV injector goes to some extent beyond proven technology.

SSC critics find the new cost estimates particularly galling in the light of the rosy picture that has so far been painted by the DoE and the project's supporters. Louise Hillsen, spokeswoman for Representative Dennis Eckart (Democrat, Ohio), an outspoken SSC opponent, says that, only six months ago, DoE told Congress that \$5,900 million was a "lean and mean figure that would stick". And after a foreboding article in the *Washington Post* in November, Representative Joe Barton (Republican, Texas), a staunch SSC advocate in whose district the project will be built, called a press conference expressly to quell fears that the price was creeping up (see *Nature* 342, 465; 1989).

Now even Barton says he needs to "look closely" at the numbers. "I've got an obligation as a responsible Member [of Congress] to make sure they have good reasons for the increase." But he is still willing to fight for a \$7,000 million SSC if necessary. "If [DoE] can really justify it, I think we can sell a \$1 billion increase", he says.

Schwitters, for his part, says that DoE never insisted that he had to build the SSC "as advertised" for the original \$5,900 million figure. "People who have experienced building things in the real world say that this [sort of cost increase] is perfectly normal", he says. But he acknowledges that coming in over budget will "fan the flames" of congressional criticism. And he says that he expects two more hard years of budget battling ahead.

G. Christopher Anderson

RESEARCH COUNCILS

New SERC chairman

London

SIR Mark Richmond is to be the new chairman of the UK Science and Engineering Research Council (SERC) from 1 October 1990 for five years, replacing Professor E. W. J. Mitchell. Richmond has been vice-chancellor, and professor of molecular biology, at the University of Manchester since 1981, and has also served as chair-



man of the Committee of Vice-Chancellors and Principals.

This will be the first time a biologist has headed SERC. Richmond's appointment may help to alleviate concern about the level of SERC spending on biological research. £28 million was spent in 1988-89 on research grants in biology, from a total budget for research and administration of nearly £400 million.

Peter Aldhous

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

The Superconducting Super Collider — how expensive?

Watkins is planning to go well armed into the budget battle with Congress.

He has convened a 'blue ribbon' panel of 15 eminent physicists, including five Nobel laureates, to give him support. By 12 January, they are expected to say that a machine downgraded to 17 or 15 TeV would not produce sufficiently good physics results to be worth building.

Watkins' prompt action came after SSC officials told him privately last month that the project could cost \$1,000 million more than the \$5,900 million that had until then been defended as both realistic and inviolable. Although earlier rumours of an impending price rise had been denied by both SSC and DoE spokesmen, SSC director Roy Schwitters now confirms that technical problems with the accelerator's magnets, among other things, will raise the project's price to "about \$7 billion". And he now says that "A machine at \$5.9 billion would be less than 15 TeV, and I don't think that's worth doing".

The single most important goal of the SSC has always been to find the Higgs boson, a hypothetical particle that plays a central role in the otherwise successful unified theory of the electromagnetic and weak forces. Most scientists feel confident

No conflict-of-interest rules

Washington

THE US National Institutes of Health (NIH) last week abandoned plans to prepare guidelines aimed at preventing conflicts of interest in NIH-sponsored research after receiving hundreds of critical comments on their proposals. In place of the guidelines, which would have required researchers to disclose in detail their personal financial interests as well as all non-federal sources of research support, NIH officials say they will propose rules that will be legally binding, but considerably smaller in scope.

Louis Sullivan, Secretary of the Department of Health and Human Services (HHS), says that "widespread concern" among researchers lay behind his decision to abandon the controversial proposals. "It is important that we do not impose on our scientific community regulatory burdens which may be unnecessary or counterproductive", he said.

The conflict-of-interest guidelines have been widely condemned by NIH-funded researchers since they were first proposed in September (see *Nature* 341, 173; 1989) after two years of pressure for regulation from the House subcommittee on human resources and intergovernmental relations, chaired by Representative Ted Weiss (Democrat, New York). Most NIH-funded institutions already have some sort of procedures to screen for conflicts of interest and researchers argue that further restrictions were unwarranted.

Because the guidelines would have applied not just to principal investigators but to "all persons who are in a position to have . . . substantive control over [NIH-funded] research", compliance would have demanded enormous quantities of paperwork.

Institutions would have to scrutinize their researchers' financial interests and outside professional activities to ensure that their work would not be compromised by, for example, ownership of stock in pharmaceutical companies. Within the University of California alone, according to its director of research and public policy, Belle Cole, meeting the requirements could cost \$10 million a year.

The guidelines might not stand up to legal assault, according to Robert Charrow of the law firm Crowell and Moring. The proposed standards would dictate policy to institutions receiving NIH funding, he says, making it arguable that they would have the *de facto* status of formal 'rules', which must go through a long approval process involving several federal agencies, including HHS and the Office of Management and Budget.

With watered-down regulations, the NIH may be able both to circumvent a lengthy approval process and to deflect

scientific protest. Current plans are to restrict the new regulations so that the total annual cost to all NIH-funded research institutions would be considerably less than \$100 million, says Katherine Bick, deputy director for extramural research at NIH.

Limiting the economic impact allows the proposed regulation to be called a 'minor rule', a bureaucratic distinction that can shorten the approval process by years by avoiding several layers of public comment. And looser regulations may be necessary to obtain the cooperation of the research community. "We want to ensure that anything of this sort is as unobtrusive as possible", she says.

Bick hopes to have proposals out by early summer. Although the comments

received in response to the now-abandoned guidelines will help form the new rules, she expects to allow for new opportunities for public comment on the issue.

According to Bick, there is only "anecdotal evidence" that a conflict-of-interest problem exists. But Weiss may not be as charitable. "[we] are concerned that NIH appears to be backing down", says James Gottlieb, a Weiss aide. If NIH officials refuse to "act responsibly" and get a grip on the conflict of interest issue, he says, Weiss may introduce legislation to do it for them. And there is evidence that the situation is not as rosy as researchers may like to believe. In a study last year by the Acadia Institute, an academic think-tank in Bar Harbor, Maine, and three other organizations, 69 per cent of the surveyed institutions reported allegations of improper conflict of interest.

G. Christopher Anderson

NEUTRINO ASTRONOMY

SNO ball starts rolling

Washington

THE Sudbury Neutrino Observatory (SNO), a \$52-million experiment that will investigate the Sun from the depths of a Canadian nickel mine, received final approval and \$30.1 million from the National Sciences and Engineering Research Council (NSERC) of Canada last week. With \$6.5 million from the project's home province of Ontario and \$14.6 million from the US Department of Energy already promised, the observatory is now waiting only for a final decision on \$1 million from the United Kingdom.

Construction will begin in the next few months, NSERC said. Located in a 10-storey cavern more than a mile underground, the facility will consist of a 14-metre-tall acrylic tank containing 1,000 tonnes of heavy water — worth \$300 million — on loan from Atomic Energy of Canada (AEC). Surrounding the water will be 2,000 photomultiplier tubes. Special low-radiation construction materials and shielding to screen out spurious particles will make the cavern the lowest radioactivity site in the world.

The huge volume of heavy water will serve as a detector of neutrinos from the core of the Sun, as well as from any fortuitous nearby supernovae which would generate a pulse of high-energy neutrinos. Neutrinos react with deuterium nuclei in the heavy water, producing high-speed electrons and positrons which generate a burst of Cerenkov radiation. The Cerenkov light is then picked up by the array of photomultipliers. Similar neutrino detectors, one at Kamioka, Japan and another under Lake Erie, detected the neutrino burst from SN1987A. But the two older detectors have chambers filled with ordinary water, whose proton nuclei offer fewer

possibilities for neutrino interaction than deuterium, and are therefore less efficient than SNO will be.

Arthur McDonald, director of SNO, hopes that the new facility will help to resolve the notorious solar neutrino problem. For many years, an underground detector at the Homestake mine in North Dakota has measured a solar neutrino flux of only about one-third that predicted by standard theory from the Sun's internal nuclear reactions. It remains unknown whether it is solar theory or neutrino physics that needs to be modified. A favoured but unproved explanation is that electron neutrinos emitted by the Sun transmute on their way here into other types, muon and tauon neutrinos, which the other detectors cannot pick up. SNO can detect all neutrino types, and so should resolve the problem. The apparatus is expected to detect several dozen neutrinos a day. But another SN1987A could generate as many as 1,000 neutrino detections in 10 seconds.

The loan of the heavy water and the donation of mine space by INCO Ltd saved the project some \$390 million, for which project planners can largely thank slower-than-expected sales of Canadian nuclear reactors. The 1,000 tonnes of heavy water comes from a stockpile AEC had intended to use in new reactors. If sales pick up, the observatory's days may be numbered; it has been promised the heavy water only for another nine years. But McDonald hopes that the basic questions on the Sun's neutrino flux will by then have been answered. After a five-year construction period, the observatory should accomplish most of its objectives within three years.

G. Christopher Anderson

AUSTRALIAN EARTHQUAKE

Twelve dead after earthquake

Sydney

THE most disastrous earthquake ever to hit Australia has left 12 dead and A\$1,500 million worth of damage. On 28 December at 10.27 a.m., the city of Newcastle, 200 km north of Sydney, experienced an earthquake measuring 5.5 on the Richter scale and lasting, according to amateur measurements, 29 seconds. The epicentre lay almost directly below the southern end of the central business district, 5 km from the centre of the city. But because the nearest scismometers are 180 km away, scientists have little to say about the geophysical nature of the seismological event.

Although earthquakes are rare in southeastern Australia, there is a recognized zone of activity extending from Victoria through most of the eastern parts of New South Wales (NSW). Newcastle lies on the northern border of the risk area identified by the Bureau of Mineral Resources (BMR). According to David Denham of BMR, south-east Australia can expect an earthquake of magnitude 5 or greater every eight years. The last earthquake in the region occurred in Wonnangatta in Victoria in 1982, and was of magnitude 5.4.

Most of the basement rocks in southeastern Australia are compressed metamorphosed sediments, granites and volcanics formed more than 400 million years ago. This is overlain by the 260 million-year-old Permian rock of the Sydney basin.

LOMA PRIETA EARTHQUAKE

Test to destruction gives clue to construction

San Francisco

RESULTS from a four-week-long series of field tests, including a simulated earthquake measuring 7.5 on the Richter scale, have indicated ways to strengthen sections of three San Francisco freeways closed since the Loma Prieta earthquake that buffeted California's north coast on 17 October, state officials announced last week. The experiments were conducted on a one-block portion of Interstate 880 in Oakland that was left standing after a mile-long neighbouring section collapsed (see *Nature* 342, 214; 1989).

Before the freeway section was demolished, hydraulic jacks applied more than 4 million pounds of force and moved it as much as 10 inches — more than during the Loma Prieta earthquake. And a pair of vibration generators fixed to the upper deck rocked the structure steadily by rotating heavy weights in opposite directions, simulating the action of an earthquake of Richter magnitude 7.5.

Investigators studied three kinds of steel bracing that could enable supporting columns of double-decked freeways to

Earthquakes usually occur deep in the basement, not in the overlying Sydney basin, and because the region lies within a single continental plate, earthquakes are not directly the result of motion at the boundary between plates. Nevertheless, says Denham, it is the build-up of pressure at the plate margins that is the ultimate

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Damaged cars in Newcastle.

cause of intraplate earthquakes. Confirmation of this explanation of the Newcastle earthquake awaits the collation of data from many seismic stations. Intraplate earthquakes, Denham adds, rarely exceed a Richter magnitude of 7; the recent San Francisco earthquake, registering 7.1, was at a well-known plate boundary.

In the past 25 years, there have been only six earthquakes in Australia that ruptured the surface. The intensity of the

withstand a major earthquake. One involved steel rods placed through the column joints. The others used steel plates and long continuous beams to bolster the sides of the columns. All proved successful in different ways. "We now have physical proof that the retrofitting concept works", said Jim Roberts, chief of the California Department of Transportation (Caltrans) Division of Structures.

The results could be important around the country: double-decked freeways similar to the Bay Area structures can be found in Los Angeles, Seattle and elsewhere, according to Jack Moehle, a University of California structural engineer who worked on the tests.

The task now is to decide which methods of retrofitting to use on five different sections of the three San Francisco freeways whose closure has created increased traffic congestion in the city.

Caltrans spokesman Jim Drago said that a decision should be made in the next few weeks, but that it will be at least six months before the freeways are repaired.

Robert Buder

INTERNATIONAL COOPERATION

Israelis and Soviets talk science

Jerusalem

IN only the second ministerial visit to Moscow since the Soviet Union severed relations with Israel in 1967, Israeli science and technology minister Ezer Weizman is this week discussing with Guri Marchuk, president of the Soviet Academy of Sciences, plans for scientific cooperation between the two countries. Weizman and Marchuk are expected to sign an agreement today, 11 January, with particular emphasis on space exploration.

Weizman's trip was almost cancelled last week when Prime Minister Yitzhak Shamir sacked him from the government for allegedly meeting, illegally, with members of the Palestine Liberation Organization. He was quickly reinstated, but is still banned from the decision-making inner cabinet.

In Jerusalem last week, meanwhile, Israel and Italy formally renewed their commitment to scientific cooperation. Officials of both countries' national councils for research and development agreed that as 1992 approaches this agreement will help Israel's chances of participating in European research projects. Lisa Perlman

Newcastle earthquake overwhelmed seismic stations in most of southeastern Australia, which were therefore unable to record the event with any accuracy. It was initially suggested that heavy mining in the area may have increased the damage caused by the earthquake, and although this idea was later discounted, disruptions to the ground structure from the mines may have caused the damage to be distributed unevenly.

The depth of the earthquake is not exactly known, but a single aftershock, felt 25 hours later and registering a magnitude of 1.5, occurred at a depth of 13 km. According to Gary Gibson, director of the Seismology Research Centre at the Phillip Institute of Technology in Melbourne, although the earthquake was deep, the local geography amplified the damage. He said that the greatest disturbance occurred in areas of alluvial fill and sediments, which amplified the seismic waves just as landfills did in the Marina district of San Francisco.

Although Newcastle has had two earthquakes of magnitude 5 or greater in the past (of 5.3 in 1868, and of 5.0 in 1925) the city is considered, by the Standards Association of Australia (SAA), to lie in a seismic zone 1, the lowest classification. But this zoning dates back to 1979, at which time the 1868 and 1925 earthquakes had not been added to the historical record. The classification is empirically based, and in fact an SAA committee met in December, a few weeks before the recent earthquake, to reconsider the ten-year-old zoning.

Tania Ewing

Oldest reptile finds loophole

London

A LEGAL loophole has opened the way to the free export of the world's earliest-known reptile fossil from Britain. The fossil was discovered in Lower Carboniferous limestones at East Kirkton, Scotland by professional fossil collector Stanley Wood, and was described in *Nature* (342, 676; 1989) and offered for sale. Wood's price of £180,000 caught UK museums by surprise, and a sale was agreed with the Museum für Naturkunde in Stuttgart, West Germany. But before the sale could go ahead, the matter was brought before the Export of Works of Art Committee (*Nature* 342, 726; 1989), which recommended a four-month delay in the grant of an export licence.

The committee's deliberations have now proved futile on two counts. First, government lawyers ruled just before Christmas that as the fossil was a geological specimen rather than an artefact, it could not be regarded as different from any common mineral and so did not fall within the committee's jurisdiction. And, second, the Stuttgart Museum says that although it still intends to purchase the fossil it will not be able to do so until the end of March at the earliest.

The delay in the sale of the fossil should allow the Royal Museum of Scotland (RMS) in Edinburgh to present the fossil as the centrepiece of its "Dinosaurs Past and Present" exhibition, starting on 1

March. The museum is also trying to raise the money to buy the fossil.

Whatever happens to the fossil, British palaeontologists are left with the unpleasant realization that "outstanding" natural history specimens are not recognized in British law and accorded the same protection as a painting or sculpture.

Beverly Halstead of Imperial College, London, who was one of the advisers to the Export of Works of Art Committee, condemns as "ludicrous" any consideration of natural objects as distinct from and inferior to works of art in their contribution to the national heritage. After all, he says, if a fossil is just a piece of rock whose importance is conditioned by our academic appreciation thereof, a painting is simply a piece of pigmented linen.

For his part, Wood sees the current legal loophole as "very welcome to a fellow like me to get the freedom to trade freely" in fossil material. Wood says he would welcome legislation that could regulate dealing in "undoubted fossil treasures" that occurs without the knowledge of national museums. He is scathing about the apparent unwillingness of UK museums to buy his wares, despite the fact that he gives them first refusal. Ian Rolfe of the RMS says that his museum was unwilling to foot the £180,000 bill until it was shown to be the market value of the unique find. The Stuttgart offer means this has now been done.

Henry Gee

Japan offers rocket power

Tokyo

IN an unusual reversal of roles, the United States may launch a satellite on a Japanese rocket in the mid-1990s. At the end of last month, the Japanese cabinet approved preliminary funds for a joint US-Japan project that would use a Japanese H-2 rocket to put into orbit a satellite to monitor rainfall in tropical regions.

The Tropical Rainfall Measuring Mission (TRMM) satellite will carry microwave radiometers, an Advanced Very High Resolution Radiometer (AVHRR) to monitor visible and infrared radiation, and cross-track scanning precipitation radar, making it a "flying rain gauge" according to NASA. The satellite and radiometers will be built by the United States and the radar and launch vehicle by Japan.

Tropical rainfall is one of the more poorly understood processes involved in determining the Earth's climate. The TRMM satellite should be able to determine 30-day tropical rainfall averages over large areas with an error of 10 per cent or less after ground-based validation, according to NASA, and will provide vital information for understanding global atmospheric circulation and ocean-atmosphere interactions such as El Niño.

The budget approved by the Japanese cabinet provides ¥58 million (\$400,000) in fiscal year 1990 for design of the interface between the satellite and rocket and between the radar and satellite. NASA began a feasibility study of the satellite last year and Katsuo Yonezawa, a senior specialist in Japan's National Space Development Agency (NASDA), is confident that both NASA and NASDA will win full support for the project.

Yonezawa says that the total cost is not clear because NASDA's H-2 rocket is still under development. But he says that Japanese newspaper estimates that it will cost each country about ¥20,000 million (\$140 million) are "reasonable". If TRMM gets the go-ahead, it will join NASDA's Advanced Earth Observation Satellite (ADEOS) in orbit. ADEOS will monitor the Earth's climate and ozone layer using sensors developed by NASA, the European Space Agency and Japan (*Nature* 335, 486; 1988).

Launch dates for both satellites are uncertain because of the difficulties NASDA is facing in developing the new H-2 rocket. During tests last summer the hydrogen turbopump of the first-stage liquid fuel engine shattered (*Nature* 340, 253; 1989) and later in the year the engine caught fire, leading to a delay of at least one year until 1993 for the first launch of the rocket.

David Swinbanks

PUBLIC HEALTH

William L. Roper named as CDC director

Washington

THE number of top-level vacancies in US public health administration was reduced by one last week when Louis W. Sullivan, Secretary of Health and Human Services (HHS), named Dr William L. Roper as director of the US Centers for Disease Control (CDC) in Atlanta, Georgia. The position had been vacant for almost a year since former director James O. Mason left CDC to become the Assistant Secretary of Health under Sullivan.

Roper, 41, will preside over a 1990 CDC budget of about \$1,000 million, of which almost half is set aside for work against the spread of AIDS. His appointment met with enthusiasm at CDC and in other areas of the public-health sector. He is no stranger to the public health service, having worked as an assistant state health officer in Alabama before taking up a series of important health-policy jobs at the White House. He will be leaving his present job as deputy assistant to the president for domestic policy on 1 March.

The directorship of CDC was one of several top health agency posts that the

Bush administration has not filled. Most notably, a director has yet to be found for the National Institutes of Health although James B. Wyngaarden left in August last year, and the departure of Frank Young two months ago as head of the Food and Drug Administration has created another conspicuous gap. June Osborn, dean of public health at the University of Michigan, says "this has been a delicate time for the agency [CDC] to be without permanent leadership", especially because of its important role dealing with the AIDS epidemic, and consequent rapid growth during the past few years.

Although she accepts the need for some "built-in deliberations", Osborn says that the waiting period has become too long. Howard Silver, director of the Consortium of Social Science Associations, believes that prospective candidates for senior positions are put off by low federal salaries, lengthy clearance procedures and ethical or political requirements but a spokesperson at HHS was quick to deny that candidates' views on ethical issues provide a litmus test for consideration.

Diane Gershon

ENVIRONMENT

Moroccan oil slick disperses

Paris

MAN and nature have finally almost dispersed a huge oil slick from the damaged Iranian tanker, *Kharg-5*, which has threatened the Atlantic coast of Morocco for the past three weeks. The 70,000 tons of light oil which spilled into

Smit Tak was refused permission to tow the tanker to shelter nearer the shore by both the Spanish and the Moroccan governments, which feared the pollution risk was too great. This, say Smit Tak and some ecology groups, such as Greenpeace, may have made matters worse,

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

The Iranian supertanker *Kharg-5* — it could all happen again.

the sea after explosions on board the tanker have now either evaporated or been dispersed with chemicals, leaving some 500 tons slowly being blown out to sea, according to the latest reports. Meanwhile the *Kharg-5* is being towed off the Canary Islands in search of shelter, with a 200-square-metre hole in her hull and 200,000 tons of oil still in her tanks.

At the end of last week, some reports from Morocco estimated the oil slick to be 175 miles long and, at 35 miles offshore, still a major threat to Moroccan beaches, sardine fisheries and oyster beds between Casablanca and Safi. But, although the Moroccan authorities now insist that there is no longer any real danger of major pollution, the coast is still under close surveillance from ocean-going tugs and light aircraft.

Morocco seems to have escaped lightly. The authorities were almost totally unprepared for an accident of this kind and the expertise and equipment needed to disperse the slick and to prevent it fouling the beaches came mostly from France and Britain. *Kharg-5* drifted for two days in heavy seas after its 32-man crew abandoned ship following explosions in one of its tanks on 19 December. During this time, an estimated 70,000 tons of light oil spilled into the sea, while the damaged tanker drifted from the site of the explosion, off the Canary Islands, towards Rabat on the Moroccan coast near the Straits of Gibraltar.

When the Moroccan authorities learned of the accident in their territorial waters, they demanded that the ship's owners, the Iranian Tanker Company, take responsibility for salvage operations. The company contracted a Dutch salvage firm, Smit Tak, which flew experts out to the *Kharg-5* on 22 December. The fire was brought under control, but heavy seas prevented either repairs to the hull or transfer of the 200,000 tons of oil to a sister ship sent by *Kharg-5*'s owners.

since the only remaining solution was to tow the *Kharg-5* further out to sea, in a storm, towards the Canary Islands some 300 miles away. The decision of the Moroccan and Spanish governments, complain Smit Tak, caused delays which allowed more oil to spill into the sea than if a speedy repair had been carried out, while leaving the threat of further pollution by *Kharg-5* still untackled.

RESEARCH IN SPACE

Juno programme outlined

London

ALTHOUGH the joint Anglo-Soviet manned space mission Juno was given the official go-ahead only in June last year, a roster of scientific experiments for the six-day mission has already been assembled. Details of the science programme were announced last week in London. Despite the compressed timescale — Juno is set for launch in April 1991 — Heinz Wolff, science director for the mission, has "no reservations whatsoever" about the standard of the 26 shortlisted microgravity experiments. A single British astronaut (Tim Mace or Helen Sharman) will spend six days aboard the Mir space station, but it is not yet clear what proportion of the planned 40 working hours will be devoted to science.

Experiments ranging from materials science to psychology are planned. In one of the largest experiments, the crystallization of up to 1,000 protein samples will be studied, following indications from experiments aboard the US space shuttle that larger and more ordered crystals may grow in low gravity. If this is the case, then a better knowledge of the structures of proteins can be obtained in a microgravity environment, but more than six days may be needed for a useful experiment to be done. Wolff is now in Moscow for several days hoping to persuade the Soviets to

According to French environment minister Brice Lalonde, who flew to Rabat on 2 January, oceanographer Jacques-Yves Cousteau first told him of the accident. The French quickly supplied chemical dispersants and a team of experts from Brest, set up after the *Amoco Cadiz* spill polluted Brittany beaches in 1978. The British sent out inflatable barrages to protect the shore — in particular the valuable oyster beds at Oualidia.

While the immediate danger to Morocco's shores has been averted, the possibility of an even greater accident, should *Kharg-5* break up, has still not been dealt with. If the Spanish and Moroccan governments do not change their minds and let the tanker enter sheltered coastal waters, it may be impossible either to repair the hole in the ship's hull or to transfer the oil to another tanker. The solution of destroying the tanker and its cargo with explosives is not winning support — both because it may fail, and because it would deprive Smit Tak of valuable salvage. The wandering tanker has also highlighted a weakness in international maritime law. No authority apparently exists to oblige *Kharg-5*'s owners to prevent a further accident while the ship is in international waters.

Peter Coles

continue with the experiment after the departure of the British astronaut.

Other experiments, involving cell culture and fungus growth, may benefit from a shorter flight time because of the difficulties of storing delicate biological material for long periods. A final list of the experiments, probably 18 in number, that will be flown on Juno will be announced later this month after Wolff's negotiations in Moscow are concluded.

The whole budget for the British side of the mission will come from commercial sponsorship. Out of a total of £16 million, only £2.3 million has been allocated to the science programme, a fraction of what is typically spent on microgravity experiments on similar flights.

Peter Graham, mission director, believes that the progress of Juno's experiments will be watched closely by the European Space Agency (ESA) and the US National Aeronautics and Space Administration (NASA) to see if this low-budget approach yields useful science.

The British National Space Centre and the Science and Engineering Research Council (SERC) will also be observing Juno with interest, following the decision by SERC in August 1989 to pull out of ESA's microgravity research programme (see *Nature* 340, 493; 1989).

Peter Aldhous

Market economy problems

London

1990 promises to be a testing time for Polish science, thanks to the country's continuing economic problems. The Polish government has already announced that there will be less central funding for research, and it is rumoured that many research institutes will be forced to close. Professor Wojtek Zakrzewski, of the University of Durham, says that Polish physicists with whom he is in contact fear for their survival.

Researchers within the university network may fare better. Andrzej Kowalczyk, of the Nicholas Copernicus University in Torun, believes that universities are protected by public and government perception of their important "social role". Certainly, there are no rumours of drastic cuts in university teaching.

The current crisis of funding for Polish science stems from the Solidarity-led government's moves towards a market economy, needed to secure vital loans from the International Monetary Fund. As part of this plan, the official exchange rate for the zloty was put within the black market range for the first time on 1 January. Although Polish scientists

should now have a freely convertible currency, the poor exchange rate (about 10,000 zlotys to the US dollar) means that purchase of Western journals and equipment may now become more difficult.

Polish science is also crippled by the bureaucratic organization of the Academy of Sciences which through its network of research institutes sponsors the major part of Poland's basic research. Set up along Soviet lines, the academy has been "top heavy . . . and dominated by the [Communist] party", according to Zbigniew Pelczynski, a fellow of Pembroke College, Oxford, and a regular visitor to Poland.

There have been changes in personnel within the academy, including the election of a new praesidium, but the stultifying structure remains. Professor Andrzej Ziabicki, on the advisory board of the Polish Society for the Advancement and Promotion of Science (TPKN), an acade-

mic pressure group closed down during martial law for "anti-socialist activities", thinks the academy cannot continue in its present shape. A more appropriate role would be as honorary body, similar to the Royal Society in Britain, he says.

TPKN has held discussion meetings on the academy's future organization, attended by government ministers, but the timing of any changes remains uncertain. The government is understood to be sympathetic to demands for reform, but the Polish parliament is snowed under with legislation, and science may take low priority.

A similar situation holds for reform of the universities. Professor Aleksander Koj, director of the Jagellonian University in Krakow, says that legislation to increase the autonomy of universities is being prepared. Ideally, this should be in place before April, when elections for university governing bodies take place. But he says it is a "good question" whether parliament will have the time to meet this deadline.

Peter Aldhous

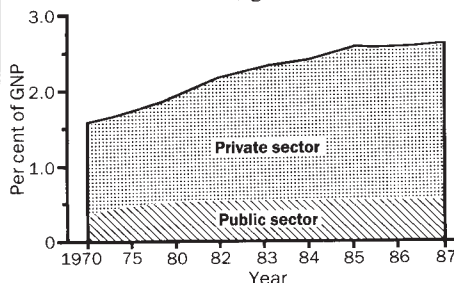
JAPAN

Bigger budgets proposed

Tokyo

JAPAN'S Science and Technology Agency (STA) has made a bold proposal that the government should double spending on research to one per cent of gross national product (GNP) in its annual white paper (policy document) released at the end of last month. This is the first time the agency has set such an ambitious target. But it remains to be seen if it will be adopted as government policy.

The white paper notes that while industry has built a powerful research and development structure through "aggressive" investment, boosting Japan's outlays for research and development to 2.6 per cent of GNP in 1988, government invest-



Where the money comes from
ment has remained flat at 0.5 per cent of GNP for the past 20 years.

Industry's investment in basic research has shown particularly rapid growth, going from ¥100,000 million in 1978 to more than ¥400,000 million in 1987 (\$2,900 million). The white paper says there is a need for the public sector to match this trend.

Yoshiro Miki, director of the policy

research division of STA, says that the agency decided to specify the goal of one per cent of GNP at this stage because the Ministry of Finance is beginning to expand the national budget after years of fiscal restraint.

But he says it will probably be "two or three years" before the Council for Science and Technology, Japan's chief science policy-making body, makes a decision on whether to adopt the agency's target.

The white paper also calls on the government to develop research facilities, such as science parks, in local regions. At present, nearly half of Japan's 215,000 corporate researchers are concentrated in the Tokyo area. Miki says that local governments have huge amounts of money which they are eager to spend on science and technology. But they lack facilities and personnel.

The Ministry of International Trade and Industry established the 'technopolis' project in the early 1980s to develop local regions (see *Nature* 329, 478; 1987). But Miki says that although the project was "far sighted" it has so far been "a failure", with one or two exceptions. Part of the problem is that, until recently, industry felt it necessary to be in the Tokyo area for reasons of efficiency. But this attitude is changing, Miki says. As an example he points to the agency's plan to establish the world's most powerful synchrotron in Hyogo Prefecture near Osaka (see *Nature* 336, 613; 1988) the operation of which will be supported by a new company set up with industry funding. David Swinbanks

Raman memorabilia gathering dust

New Delhi

As India celebrates the centenary of the birth of Nobel laureate physicist C. V. Raman, his university, Rajabazar College in Calcutta, has been embarrassed by the revelation that a collection of musical and scientific instruments used by Raman is mouldering in a college classroom. Lack of space and money, the college says, has meant that the historically important instruments have been left unguarded and unprotected.

The story goes that the discovery for which the physicist won the Nobel prize — the Raman effect in spectroscopy — was made with equipment worth about 100 rupees (\$6). In later years, Raman devoted his energies to acoustical studies, especially of musical instruments, and acquired a piano and an organ. According to a college spokesman, these instruments, now rusty, were left in an open classroom to provide a source of inspiration to students.

Although these same students are now complaining that the college has been negligent in its upkeep of Raman's memorabilia, the college is resisting efforts to remove them to a museum. But it says that it would like to preserve the items "in a better way" on the college campus, if money and a suitable location can be found.

K. S. Jayaraman

Surveying sexual attitudes

SIR—S. J. Gurman writes to discuss the problem of obtaining valid data on sexual attitudes and lifestyles through survey methods (*Nature* 342, 12; 1989). He uses an analogy from natural science to make the point that the mean number of heterosexual partners should be equal for men and women. In mathematical models of the transmissions dynamics of sexually transmitted diseases, this constraint plays a central role in model formulation and analysis (see *AIDS* 3, 333–346; 1989). However, the constraint mentioned by Gurman holds only for a closed population without immigration and emigration throughout a time period in excess of the maximum life span of sexual activity. In our oral presentation of the data, on which the News and Views article in *Nature* was based (*Nature* 341, 181; 1989), we were at pains to point out the difficulties inherent in interpretation (incidentally, the mean number of 'lifetime partners' was misreported in the article; our figures show a smaller disparity between means for men and women).

Our data show high variability and a very skewed distribution in reported number of sexual partners. Not surprisingly, there is less variability in numbers of partners reported over short time periods, for example one month, but this increases markedly with the length of time interval for which numbers of sexual partners are reported (up to "in your lifetime (so far)").

The estimates of the mean lifetime partners for men are heavily influenced by the very small proportion (less than 1 per cent) reporting high numbers of partners (>100). Of equal interest is the overall pattern of distribution of numbers of partners. Sample bias is of particular concern in the estimation of the population mean or variance — variance is related to the mean by a power law relationship, with a power in excess of 3 (*Nature* 333, 514–519; 1988). In order adequately to represent the variability in sexual behaviour, sample size, as well as selection, is important. In the feasibility study, the sample size was approximately 1,000, in the full study it will be around 20,000.

Both increasing time and increasing numbers of partners lead to greater difficulties of recall, and therefore estimates of lifetime partners are inevitably the most unstable. We found much greater similarity between men and women in patterns and numbers of partners when shorter finite time intervals were considered.

'Lifetime partner' estimates for each individual do not refer to a finite time period. They include raw data for all ages and refer to partners since first intercourse to the present time. They therefore report numbers of partners 'thus far'. These raw

data take no account of age structure of the sample, age mixing, generational and demographic changes, or international mobility of populations, all of which affect patterns of sexual behaviour. Our data show that men choose female partners who are, on average, younger than themselves. Thus the age distribution of women in our sample is not necessarily representative of the age distribution of the women partners of the men in this sample (and vice versa).

The quantitative study of human sexual relationships is as yet a little-explored area of research and the methodology is still being developed. In interpreting the data, caution will be uppermost in our minds. While the analogy to be drawn between the behaviour of chemical bonds and human partnerships is useful to evaluate data on patterns of sexual behaviour, there is a complex set of factors influencing patterns of sexual behaviour and mixing on which we hope that our main study will shed further light.

A. M. JOHNSON

University College and Middlesex
School of Medicine,
London W1, UK

J. WADSWORTH

St Mary's Hospital Medical School,
London W2, UK

J. FIELD

Social & Community Planning Research,
London EC1, UK

K. WELLINGS

Health Education Authority,
London W1, UK

R. M. ANDERSON

Imperial College of Science
& Technology,
London SW7, UK

Fruits of the forest

SIR—In their Commentary article (*Nature* 339, 655–656; 1989), Peters and colleagues show that exploitation of non-wood resources in the Amazon rainforest, especially fruits, is far more economically rewarding than timber extraction. Therefore, sustainable use of these resources might provide an enticing argument against timber operations and an important step towards conserving the Amazon.

Current methods of harvesting many Amazonian fruits are alarmingly destructive, however. The most valuable fruiting trees, the palms, are cut down and killed in order to collect the fruit, because their physical structure makes climbing virtually impossible. Palm trees that are excessively harvested in this manner include *Mauritia flexuosa*, *Mauritiella peruviana* and *Jessenia bataua*.

Our studies, yet to be published, show that local inhabitants in Peru use palm

fruits in a non-renewable manner that may be more detrimental than selective timbering. This is because large numbers of fruiting trees are permanently removed from the rainforest ecosystem just as they are beginning to be productive, whereas timber trees are normally cut in the non-fruiting state. Local inhabitants will not cut palm trees that grow on their small private plots, but will collect the fallen fruits and use them for household consumption. It is fruit from the vast areas of government-owned forests that are collected for commercial sale, and these trees that are cut during the harvest.

Cutting of palm trees also decreases the populations of ungulates and large rodents that rely on palm fruits for food. Reduced mammal populations result in diminishing returns of game meat for subsistence hunters and lower protein intake for local people. Unfortunately, until alternative non-destructive methods of harvesting fruits can be developed, there is no basis for arguing for fruit rather than timber as an economic resource in the Amazon.

RICHARD E. BODMER

TULA G. FANG

Large Animal Research Group,
Department of Zoology,
University of Cambridge,
Cambridge CB3 0DT, UK

LUIS MOYA

Ministerio de Agricultura,
Region Agraria XXII-Loreto, Apt 621,
Iquitos, Loreto, Peru

MIT in the red

SIR—I write to clear up two misstatements referring to me in Seth Shulman's News report "MIT in the red" (*Nature* 341, 558; 1989).

To begin with, I may be thought of as a critic of excessive or misguided US strategic nuclear weapons procurements, but not of "MIT's reliance on military research dollars". I know of no evidence that Department of Defense funding has distorted academic research at Massachusetts Institute of Technology (MIT), and I have never said that.

Second, the "high level" meeting Shulman refers to has been called for the purpose of exploring ways to redeploy to the civilian sector some of the US research and development resources now used in weapons' development, in order to create an environmental technology base and to improve US industrial competitiveness. The meeting will be held at the American Academy of Arts and Sciences, and has nothing whatsoever to do with research at MIT.

KOSTA TSIPIS

Program in Science and Technology for
International Security,
Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, USA

Publishing and perishing

SIR—In his article "Publishing and perishing" (*Nature* **341**, 349–350; 1989) John Merriman does a masterly job in describing the serials crisis faced by academic libraries. I dispute, however, his conclusion that commercial publishers are not a major cause of this crisis.

Using figures from Blackwells the subscription agents, he says, "For 1989 over 1988 . . . commercial publishers increased their [journal] subscription prices by 10.57 per cent, the learned societies by 10.40 per cent. Nor are overall prices notably different — £207.50 for commercial publishers, £217.85 for learned societies". Yet there is abundant evidence showing that less expensive journals tend to originate with societies rather than commercial publishers. Such findings are based on analysis of price per character, word, 1,000 characters, or other normalized measure, and reveal that in certain subjects, societies and university presses generally publish information at less than commercial companies, and sometimes as little as 5 per cent of commercial prices.

An overview of these studies, which establish a more general pattern, is to be found in the *Report of the ARL Serials Prices Project* (Association of Research Libraries, Washington, DC, 1989).

DUANE E. WEBSTER

Association of Research Libraries,
1527 New Hampshire Avenue, NW,
Washington, DC 20036, USA

Citing patents

SIR—The comments from Frank W. Cousins (*Nature* **341**, 380; 1989) on the necessity of citing patent numbers are useful because the patent literature is important. Each year, more than a million patent publications are produced worldwide. But the patent number is not sufficient in many cases. Many patent offices use the same numerical sequences; many begin a new series each year. A number such as 88/1234 could be a patent number in many different countries. So it is always necessary to add the country code of the issuing patent office. These two-letter codes (for example GB for United Kingdom, US for USA, DE for Bundesrepublik Deutschland, DD for DDR) are concise and are available even for very small countries such as Vanuatu or Tuvalu. The codes are standardized in the international standard ISO 3166 which is frequently updated. They are preferable to automobile country codes (which do not exist for all countries).

Unfortunately, country code plus patent number is still not sufficient for a distinctive citation. The same number is

used by many patent offices for various procedural stages (for the same invention). The resulting publications are usually not identical. To give an example: for DE 2338518, three different publications have appeared for the patent specification as filed (open to public inspection, then the examined and accepted specification and the granted patent). To differentiate between them, a document code is necessary which should be added to the number to make it unique (for example DE 2338518-C for the granted version).

A kind of document code nowadays is very often given on the printed specification. There will be a universal standard by WIPO (the World Industrial Property Organization in Geneva). This is especially important if one orders a Japanese publication. The same number (with different kinds of document codes) is used for at least four completely different specifications which do not belong to the same invention.

Patents are not cited very frequently in the journal literature although they may be the only source for technical innovations or at least the earliest one. Many people probably fail to cite them because they do not know how to do so.

A good tool for citing both normal and exotic types of publications is the German standard DIN 1505 part 2 (Zitierregeln; Bibliographic references to documents; rules for citing). Four-and-a-half pages are devoted to patent citations, but there are other examples for citing nearly everything, even a 78 r.p.m. jazz record or something from a museum collection.

MARTIN A. LOBECK

Strundenerstrasse 100,
D-5000 Köln 80, FRG

Separate culture

SIR—You are correct (*Nature* **340**, 413; 1989) in saying that the black and coloured people of South Africa "have other things on their minds" than English sports, such as cricket, rugby football and tennis. Their interests are soccer, boxing and athletics, and, for the younger generation, also education, the subject of the Commentary in the same issue of *Nature* (page 420).

I had the opportunity to see and learn to know South Africa first as an immigrant, later as a citizen and then also from abroad. South African blacks, unlike American blacks, still live in their natural geographical regions and surroundings. They are divided into nine major and several smaller tribes, each with its own culture, tradition, habits and language. These circumstances are usually unknown to, or ignored by, foreigners to South Africa.

For at least 80 per cent of all black scholars, first contact with the English and Afrikaans languages (the languages of instruction) is at school. Separate

school education, at least at primary level, is thus necessary. That is a cultural issue, not a political issue stemming from *apartheid*.

Because of the increase in the black population, more schools have to be built in these geographical regions. Black students cannot attend white schools, which may be more than walking distance away, so that schools have to be brought to the scholars because of the lack of transport systems. And transport systems cannot be created from taxes, paid by the already heavily burdened small percentage of taxpayers.

As to higher education, any student with the necessary exemptions in the matriculation examinations has the opportunity to enrol for higher education. As correctly mentioned in the two *Nature* articles, enrolment is possible at any university. The reasons for the low proportions of black students thus have to be found elsewhere, and are not political. South Africa is concerned about the figures, but is not the only one thus affected. Similar figures for black student enrolment apply in the United States (*Science* **244**, 1536; 1989).

G. OSTHOFF

Department of Biochemistry,
Bowman Gray School of Medicine,
Wake Forest University,
Winston Salem,
North Carolina 27103, USA

Brazilian research

SIR—In your extensive review of science in Brazil (*Nature* **342**, 353–374; 1989), we find it most surprising that so little mention is made of the distinguished research on tropical diseases there, other than that at the Instituto Oswaldo Cruz in Rio de Janeiro. Two laboratories with considerable international reputations in this field are the Instituto René Rachou in Belo Horizonte, notable particularly for its contributions to schistosomiasis and malaria, and the Instituto Evandro Chagas (IEC) in Belém for its work on malaria, leishmaniasis, hepatitis and diarrhoeal diseases.

At the IEC, important studies on arthropod-borne viruses were carried out in the 1950s and 1960s, with Rockefeller Foundation support; more recently, there has been remarkable work there on the epidemiology of leishmaniasis of the Amazon region by Ralph Lainson, Jeffrey Shaw and their colleagues, generously supported by the Wellcome Trust for the past 20 years.

DAVID WALLIKER

Department of Genetics,
University of Edinburgh,
Edinburgh EH9 3JN, UK

GEOFFREY TARGETT

London School of Hygiene
and Tropical Medicine,
London WC1 7HT, UK

The grant-getting game in Japan

Mitsuhiro Yanagida

Biologists in Japanese universities are hamstrung by an obscure and unfair system of allocating grants. Reform is needed if Japan's scientists are to compete with first-class international research groups.

In Commentary articles published last year, Akio Yamamoto explained how Japanese research is only meagrely supported by the government¹ and Shahid Siddiqui argued that foreign nationals are not encouraged to work in Japanese universities². Here, I wish to draw attention to another important issue — the system of providing grants for biology research in Japanese universities.

Contrary to general belief, many laboratories in Japanese universities are shabby and poorly equipped. Research of the highest quality is still produced, but if university laboratories were allowed more funds, and more flexible use of them, the development of bioscience in Japan would be greatly accelerated. The amount of money needed would not be great compared with the size of the country's economy. The main problem university researchers experience is unpredictability and difficulty in acquiring research funds. These problems also occur in the United States and Europe, but to my mind they are far more severe in Japan.

To run a medium-to-large-sized laboratory such as my own, where three staff scientists, one English postdoctoral fellow and twelve graduate students are working on the cell and molecular biology of yeast, an annual budget of at least ¥20 million (\$140,000) is required. The only way this can be achieved on a single grant is by obtaining a 'special distinguished' award (see below). Otherwise, the money has to be collected as five or six small grants for different projects. Writing proposals and reports and attending conferences for each grant is time-consuming — but, of course, far better than having nothing.

Government support

There are five government agencies providing support for the biological sciences: the Ministry of Education, Science and Culture (MESC, also known as Monbusho) and the Science and Technology Agency (STA); and the ministries of Agriculture, Forestry and Fisheries (MAFF), Health and Welfare (MHW), and International Trade and Industry (MITI). There is no central organization equivalent to the National Institutes of Health or the National Science Foundation in the United States. The Japanese governmental agencies have very different policies. MESC mainly finances university research, MITI funds industry and STA supports both public and private research.

It is surprising that, given all of these agencies, there is only one category of fund, from MESC, openly available to researchers in universities. To obtain support from other government sources, researchers have to wait to be asked.

Traditionally, university researchers have lived under the financial umbrella of MESC, which appoints all staff in public universities. The other ministries are hesitant or unenthusiastic about supporting university researchers who 'belong' to MESC. This custom is becoming increasingly restrictive as biological research becomes more expensive. Other government agencies now have large sums of money for bioscience, medical research and biotechnology, and want university researchers to participate in their projects (to ensure international collaboration, for example). But universities are not free agents, and relations between MESC and the other agencies are often tense.

Scientists in Japanese universities can apply for MESC's Grant in Aid for Scientific Research (*kakenhi*) only once a year, with a chance of success of about one in three this year. Many grants are already reserved, because a researcher has to reapply for the same grant every year, even if it has been awarded for three years. The real chance of success is therefore one in five or less. Unsuccessful applicants receive no information as to why a proposal has been rejected.

The total amount of *kakenhi* was ¥526 hundred million (\$365 million) for 1989. Even though the budget has doubled during the past ten years, this is not a large sum of money because it has to cover all of science (including the natural, human and social sciences, and engineering, medicine and agriculture). There were 61,000 applications for 1989, of which 19,000 were accepted, an average of ¥2.8 million for each approved grant. Funds have to be spent on equipment and materials, and cannot be used for salaries. The total sum of *kakenhi* is indeed small, given the size of Japan's economy: it is only 10 per cent of Toyota's annual profits, for example, or 15 per cent of the annual budget for research and development in an electronics company.

Most researchers have no *kakenhi* grant; some do not even bother to apply for it as the competition is too great. So how can they carry out research? Staff in public universities automatically receive a fixed award from MESC each year. A

large part of this budget is needed for maintaining essential services such as electricity, gas, water and telephones. The amount of money left over varies, depending on the university or department, though it generally ranges between ¥0.5 and ¥1.0 million per member of faculty. This is not much, but it is crucial for researchers who are not supported from other sources. I do not know what percentage of scientists in public universities have to rely on this budget alone, but it is likely to be high considering the high rejection rates of *kakenhi*.

Freedom of choice

Despite its serious defects, *kakenhi* at least has an open application system and allows researchers freedom to choose projects. There are basically two classes of grant that individuals can apply for: 'special distinguished' and 'general'. The first class (¥50–300 million each for a period of three to five years) is awarded to internationally recognized researchers in bioscience. Last year there were 24 qualified applications in the entire field of bioscience, only four of which were accepted. The second, 'general' class is subdivided into three categories (A, B and C, worth ¥10–50, 3–10 and less than 3 million, respectively). This year there were 488 and 3,236 applications for A and B grants in bioscience, respectively, only 85 and 504 of which were accepted (rejection rates of 83 and 84 per cent). The B grants, which are awarded for projects running for two to three years, may be sufficient only for a one-year budget for a small laboratory, and C grants are suitable only for individual investigators. The A grants are obviously the most attractive, but the chance of getting one is likely to occur only once in a lifetime, while the high rejection rates and insufficient funding of the 'general' category make the life of laboratory chiefs in Japanese universities difficult.

Other types of *kakenhi* grant are given to specified priority areas such as cancer research. These are perhaps characteristic of Japanese science and technology, as most of these grants go to teams of scientists, the individual members of which are predetermined and work under a chief investigator. If researchers want a secure source of income, it is vital to become a member of one of these teams.

'Priority' grants usually run for three years, and every year six to nine new areas are selected from many applications (the

number of applications is not disclosed, but the rejection rate is probably greater than 85 per cent). An MESC committee of university presidents, professors and scientists selects the fields. In 1989, these were AIDS, Alzheimer's disease, animal development, the bacterial genome, the genetic code, cell motility, nerve-cell impulse and signalling, and plant reproduction. Each of these areas is earmarked to receive ¥100–200 million a year, and the projects are usually run by three or five teams each consisting of 20–50 members. 'Priority' fields usually end within three years, and researchers then have to compete for the new 'priority' areas.

Individual proposals for the 'general' category of *kakenhi* are reviewed by a small number of committee members selected by academic societies or by MESC, but the system is very different from that of peer-review practised in the United States and elsewhere. The number of reviewers for each field is so small (usually three) that each reviewer has to cover topics outside his or her area of professional expertise. The reviewers rank each proposal into five classes, but are not obliged to justify their opinion, so the reviewing system is basically a black box.

In applying for 'priority' grants, a group of researchers submits to MESC a written proposal describing the need for the research proposed, the importance of the area, the group's plan and a budget, together with the list of participating scientists. If this succeeds, the next step is an oral presentation by the principal investigator to MESC's reviewing committee for all bioscience research. Because of the high rejection rate, the atmosphere of the examining committee is tense, both for applicants and perhaps also for committee members. I spent two years' effort in proposing a project on the molecular biology of chromosome structure and behaviour, one of the 'priority' fields for 1990, and it was very frustrating that no reason was given for rejection. The outcome is very unpredictable for applicants and a great deal depends on who decides on the composition of the committee.

Other sources of funds for university researchers are private foundations and industry. Most foundations encourage applications and play a large part in supporting research. Many good laboratories are supported by well-known foundations such as Yamada, Mitsubishi, Toray and Nissan. The total sum of foundation grants available for bioscience, however, is tiny compared with that in the United States. It is impossible to obtain research grants from private companies in Japan unless the individual has a special relationship with the company, such as a research contract, training company researchers in the laboratory, or sending graduates to

the company. Figures are not available for how much company money has been put into research in university laboratories. But the distribution must be very uneven, and is almost certainly concentrated in particular areas of medical research and biotechnology. Certain clinical medicine and pharmacology laboratories are thought to receive large grants from pharmaceutical companies. In most other departments, such as ours, funds from companies are negligible, one reason being that there is no tax relief for the donation of funds to universities in Japan.

Japan was once notorious for cheap labour. But the annual income of university professors is now similar to that of their counterparts in the United States, although the sums involved are probably not comparable because of the high cost of living in Japan. Furthermore, research in Japanese universities is largely supported by graduate students who are paid little. In public universities these students have to pay fees of ¥320,000 per year, although they can receive a fellowship (approximately ¥80,000 per month) from Japan Ikueikai Foundation. This fellowship can, however, turn out to be a burden. If a student gets a job in industry or goes abroad after qualifying, he has to pay back the money (plus interest in some cases).

The problems for biological researchers in Japan can be summarized as follows. Although increasing amounts of money are flowing into bioscience and biotechnology, only a very small fraction of that money is openly available to researchers in universities. Many government funds are targeted and distributed to industry or non-university institutions without open application and rigorous peer review. Even in MESC's *kakenhi* system, the reviewing process is far from ideal. There is no organization of scientists capable of changing the technology-orientated attitude of government agencies and of being involved in distributing funds among the agencies. The Japan Academy hardly plays any role in representing scientists in congress or government.

How can these problems be tackled? As a first step, rejected *kakenhi* proposals should be returned with comments by (anonymous) reviewers. This would surely improve the general feeling among scientists about *kakenhi* and also, in time, improve the quality of scientific research. More travel expenses for members of the reviewing committee (like the US study section) would be worth the investment. For larger grants, such as the 'special distinguished' and 'general A' awards, foreign scientists should be included on the reviewing committees, as real scientific peers are rarely found in Japan. There should be no serious problem in writing proposals in English for those applicants whose work is supposed to be 'internationally well recognized'. It would be

helpful if grants could be applied for two or three times a year. And graduate students should not have to return their fellowship awards if they work in non-profit organizations in foreign countries.

It would be a significant improvement for university researchers if they could apply for some of the grants from government agencies other than MESC. STA, MAFF and MITI grants often run for five years, for example, and a single grant is generally larger than those offered by MESC. Cases such as one agency supporting obscure research groups in industrial companies with roughly ¥100 million per year each for five years, indicate that a redistribution of government money for bioscience and biotechnology is indeed necessary. Evidence that the situation can be improved is provided by the Protein Engineering Research Institute, recently founded by MITI and the private sector, which is already held in high regard by the academic world. The Human Frontier Science Program is a promising venture that will support the research activities of international teams in the areas of brain and molecular biology research. In Japan, the task force for the Human Frontier Program is decided by STA; for the future it would be better to create a new foundation, equivalent to the US National Science Foundation, which would distribute government money for basic research.

Only a radical increase in the total amount of *kakenhi* or other accessible grants will lessen the frustration that has built up among researchers in Japanese universities. Asking for a tripling of grant money may be unrealistic, but I think that this is what is needed if an attempt is to be made to gain a competitive edge with first-class laboratories in the United States and Europe. Furthermore, the money should be used for paying the salaries of researchers from any country.

The increase of funds itself will not solve other deep-rooted problems in the Japanese scientific community, but it is a prerequisite for improvement. I hope in particular that greater opportunities will be given to young researchers, and that universities will look for money from the private sector at the same time as appealing to the government for tax benefits for those investing in university research. □

Mitsuhiro Yanagida is in the Department of Biophysics, Faculty of Science, Kyoto University, Sakyo-ku, Kyoto 606, Japan.

1. Yamamoto, A. *Nature* **339**, 575–576 (1989).
2. Siddiqui, S. S. *Nature* **340**, 337–338 (1989).

Erratum: drug trials in Japan

In the Commentary article by Masanori Fukushima (*Nature* **342**, 850–851; 1989) the figures for disqualification rates in the clinical trials discussed should have read "18 to 26 per cent" (not "18 to 36 per cent"). Later, "multiple melamona [sic]" should have read "multiple myeloma" in the passage describing the approval of interferon- α .

Publishing without being damned

A remarkable research report claims that an object spinning on a vertical axis weighs less if spinning clockwise (seen from above) than anticlockwise, but it may be too soon to cast off Newton's laws.

WHAT should a journal do when a contributor sends it a piece of research whose conclusions fly in the face of current beliefs? There are several responses, among which the exercise of studied scepticism is the most common. Contributors can be challenged almost indefinitely by increasingly zealous referees and may sometimes throw in the sponge. But what happens if the referees find nothing wrong, but are still unpersuaded of the conclusions? In a celebrated case in 1988, this journal published such a research report, following it by an on-site investigation and a denunciatory report, which concatenation of events was itself a cause of controversy.

In very different circumstances, *Physical Review Letters*, the American Institute of Physics' most eagerly read journal, has just dealt with such a contribution much more simply, merely by publishing it. No disrespect to any of those concerned, authors, referees or editors, is intended by the remark that the outcome is thoroughly unsatisfactory for the other interest-group vividly engaged in the publication process — the readers.

These are the circumstances. Two Japanese at the Tohoku University in Sendai, Hideo Hayasaka and Sakae Takeuchi, have carried out a simple experiment with a gyroscope (or, more exactly, with three gyroscopes): they have simply set it spinning and then weighed it, using a chemical balance (*Phys. Rev. Lett.* **63**, 2701; 1989). Some measure of the journal's embarrassment can be gleaned from the time spent in brooding on this research report, which was received for publication on 7 March 1988, revised by the authors by 9 August last year and published only at the beginning of the Christmas torpor.

The expectation is, of course, that the weight of a spinning gyroscope should be independent of its rotational speed or of whether or not it is spinning at all. What the authors find is that expectation is confirmed if the angular momentum vector of the rotation is directed upwards, against the direction of the Earth's gravitational field, or if the rotation viewed from above is anticlockwise. But — this is the surprise — if the angular momentum vector is directed downwards, towards the centre of the Earth, the weight is decreased.

The effect seems unambiguous, amounting to as much as 10 mg in a total of 175 g (the weight of the heaviest gyro-

scope), which is well within the sensitivity of the balance (0.3 mg). The reduction of weight increases linearly with rotation speed and, for a given speed, is the same fraction of the weight. The balance measurements have been confirmed with a quite different electronic instrument in which deflections of a horizontal lever are compensated for by electromagnetic forces.

In all the measurements, the rotor of the gyroscope, housed in a metal cage, has been set spinning by means of a voltage amplifier, which is then disconnected. For the purposes of measurement, the gyroscopes have been placed within evacuated chambers sitting on one balance pan. In the measurements with the chemical balance, made of non-magnetic materials, the weight changes have been recorded both while the gyroscopes are shielded locally and then when the whole equipment is placed within a magnetically shielded room. Dutifully, but unsurprisingly, the authors have turned their spinning gyroscopes upside down, verifying in the process that the asymmetry they observe is simply a function of the direction of rotation, not of the orientation of the gyroscopes. The only obvious precaution against systematic errors not mentioned is that of swapping the gyroscopes from one balance pan to the other, but it is difficult to see why that should make a difference to the outcome.

Needless to say, there is no way in which the observations can be accommodated within conventional theories of mechanics, classical or otherwise. Indeed, the authors state laconically that "the experimental result cannot be explained by the usual theories". The best bet is some kind of spin-spin coupling — in this case between the rotation of the gyroscope and that of the Earth — such as would be allowed by some versions of Einstein's theory of gravitation, but the measurements are uncomfortably huge by those standards and, in any case, there is no reason why parallel spins should repel each other when antiparallel spins are indifferent to each other's presence.

So what is to be made of the result? Curiously, what might justly be considered the most radical announcement to touch basic physics since Einstein's time, the fifth force of the past few years notwithstanding, has so far occasioned very little comment. Those convinced that the general press is forever seeking to publi-

cize sensations will remark that newspapers and other general media have shied away from a story that would have justified some delicious headlines. Why should that have been the case? Can it be that the scientific community itself is discomfited? Certainly this journal has found opinions of the work of Hayasaka and Takeuchi sparse in the past few weeks. (Even so, one courageous spirit promises a publishable opinion very shortly.)

The plain but conventional truth is that it is too soon to take a view. On that view, the results, correct or not, are interesting but inconclusive on their own. The next step will be to see whether they are confirmed or, alternatively, contradicted by independent measurements. Then, the argument goes, will be time enough to wonder what, if anything, they mean. Meanwhile, the small but zealous band of those devoted to the notion that suitably managed gyroscopes contain a recipe for the levitation of massive objects and even, perhaps, for perpetual motion can be counted on to fill the months ahead with endless noise. No doubt, theoreticians skilled in the ways of chirally asymmetric particles such as neutrinos will join in the fun.

That is why it would have been helpful if Hayasaka and Takeuchi had said a little more about their measurements. How, for example, did they come to be weighing a rotating gyroscope? "To confirm the reflection symmetry relating to the rotational motion of objects in the gravitational field of the Earth . . ." is what they say, but had they cause to believe that this point had not previously been tested? With all those spinning satellites in orbit about the Earth accumulating upwards kicks of acceleration whenever their spin axes coincide with the upward vertical?

There are also pernicky questions to ask about the data — why are the error bars for weight, described as "the fluctuation of weight changes", greater for clockwise than for anticlockwise rotations at the same speed? No doubt all that will in due course be made plain, but how will Hayasaka and Takeuchi occupy themselves in the meantime? The common *canard* that the science establishment (of which *Physical Review Letters* is certainly a part) is forever suppressing uncomfortable truths has been shown to be false, but those who astonish us in ways like these have the first duty to check that what they have to say stands up.

John Maddox

Pitch perception in silicon

Brian C. J. Moore

MANY commonly occurring sounds, such as those produced by musical instruments or the human voice, are perceived as having pitch. The pitch heard generally corresponds to the fundamental frequency of the sound, even when the fundamental component is missing or masked by noise. Humans are remarkably good at extracting the pitches of sounds under difficult listening conditions, such as when reverberation or background noise are present, and it has proved exceedingly difficult to produce electronic devices or computer algorithms that come close to human performance in this respect. The mechanisms by which pitch is extracted have been the subject of intensive study over the past 150 years¹⁻³, and although some uncertainties remain, much is understood.

Now Lazzaro and Mead⁴ have designed an integrated circuit (hereafter called the chip) which models many of the mechanisms believed to be involved in human pitch perception. The 125,000-transistor chip uses analogue processing and runs

in real time. It has the potential both of allowing the evaluation of models of pitch perception, and of being used as a practical and robust pitch-extraction device in speech-recognition systems, speech-pattern hearing aids for people whose hearing is severely impaired⁵, and cochlear implants⁶.

Figure 1 is a block diagram of a model of human pitch perception³. In the first stage, the signal passes through an array of bandpass filters which perform a crude spectral analysis of the sound. These filters, known as the auditory filters or critical bands, have their physiological basis in the filtering action of the basilar membrane within the cochlea. The chip includes 62 such filters. Figure 2 shows a simulation of the outputs of 17 auditory filters, with centre frequencies (CFs) ranging from 159 to 6,400 Hz, in response to a periodic complex sound (a pulse train with a repetition rate of 200 pulses per second). At low CFs the filters resolve the individual harmonics of the complex sound, and the outputs of the filters are almost sinusoidal. At high CFs, however, the filters are too broadly tuned to resolve individual harmonics, and the outputs are complex waveforms whose repetition rate equals that of the input waveform.

The next stage in the model is neural transduction; the analogue waveforms at the outputs of the filters are transformed into discrete neural impulses. Circuits in the chip perform operations similar to those known to occur in the hair cells of the cochlea, namely half-wave rectification and nonlinear amplitude compression. These are followed by circuits modelling the neurons in the spiral ganglion within the cochlea. The output of each model neuron is a sequence of pulses, analogous to neural action potentials or 'spikes', whose probability of occurrence is greatest near peaks in the fine structure of the waveform at the output of the filter which 'drives' that neuron.

Next comes the extraction of information about the timing of neural impulses. There is considerable uncertainty about how this is done in the human auditory system. Lazzaro and Mead have followed a suggestion of Licklider⁷ and use circuits which compute a running autocorrelation in each channel. This is accomplished using an array of delay lines and logical AND elements. The autocorrelation function in each channel shows peaks at times corresponding to prominent periodicities in that channel, and at integer multiples of those times. For example, in Fig. 2 the output of the filter centred at 800 Hz is

roughly an 800 Hz sinusoid, and the autocorrelation function for that channel would have peaks at 1.25, 2.5, 3.75, 5 milliseconds and so on.

The following stage combines the time-interval information across different CFs. Lazzaro and Mead have implemented this simply by summing the autocorrelation functions across channels. A single wire running vertically through the chip acts as a dendritic tree to perform the summation for each time delay. In general, the largest peak in the summed autocorrelation function gives a time interval corresponding to the perceived pitch. For the stimulus in Fig. 2, that interval is 5 milliseconds.

Lazzaro and Mead have shown that their chip extracts pitches close to those perceived by human listeners for many of the stimuli used in classical studies of pitch perception, including non-harmonic complex tones and comb-filtered noises. But the chip does show some systematic deviations from human perception, traceable to the following factors. First, the filtering in the cochlea is known to be nonlinear, whereas in the chip it is modelled by linear filters. In response to complex

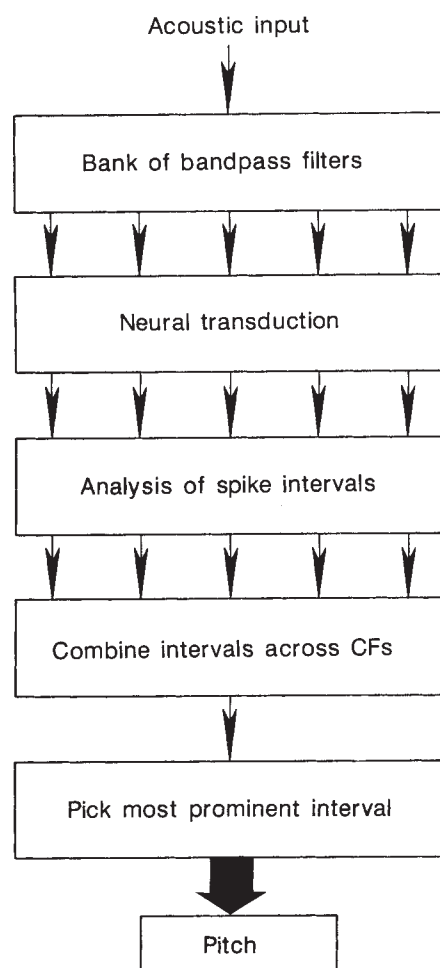


FIG. 1 A schematic model for the perception of the pitch of complex sounds.

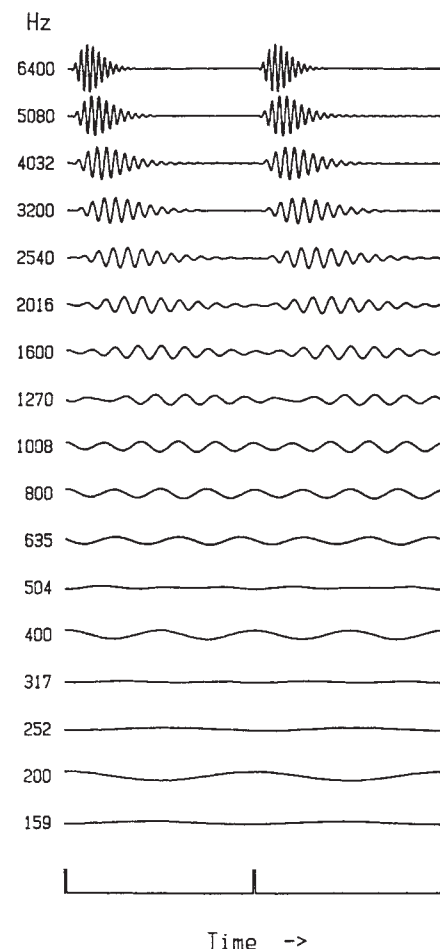


FIG. 2 A simulation of the outputs of 17 auditory filters in response to periodic impulses with a rate of 200 pulses per second. The centre frequency of each filter is indicated on the left. The stimulus waveform is shown at the bottom.

GREENHOUSE EFFECT

Reduced rise in sea level

M. F. Meier

tones containing only a few high harmonics, the cochlea's nonlinearity results in combination tones that act like lower harmonics. These combination tones can have a considerable influence on perceived pitch⁸, but they are not modelled by the chip. A second problem is that all of the chip's channels are weighted equally in the computation of pitch. In human pitch perception, however, the channels responding to the lower harmonics appear to receive more weight than the channels responding to the higher harmonics³. A third drawback is that the chip produces only a single pitch value for a given sound; this is achieved using a 'winner-takes-all' nonlinear circuit. For humans, however, some sounds may be perceived as having more than one pitch. Indeed, this ambiguity of pitch has had an important influence on the development of pitch theories.

In spite of these limitations, the chip represents a major advance. Other researchers have implemented computational models of pitch^{9,10}, but these models do not run in real time and they require substantial computing power. The chip is compact and runs in real time. Like neural structures, it performs robustly despite random variations in the components making up the circuits. Indeed, Lazzaro and Mead argue that this is an important property of integrated-circuit models of neural structures. Although they do not present any data on how well the chip can extract pitch in noisy situations, it should perform rather well if it captures the essential features of human pitch perception. It can therefore be expected to have application in devices that require real-time, robust pitch extraction, such as signal-processing hearing aids⁵ and cochlear implants⁶. But it remains to be seen whether the performance of the chip in noise is superior to that achieved using 'multilayer perceptron' pitch extractors¹¹. □

Brian C. J. Moore is in the Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK.

1. Ohm, G.S. *Ann. phys. Chem.* **59**, 513–565 (1843).
2. von Helmholtz, H.L.F. *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* (Vieweg, Braunschweig, 1863).
3. Moore, B.C.J. *An Introduction to the Psychology of Hearing*, 3rd edn (Academic, London, 1989).
4. Lazzaro, J. & Mead, C. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9597–9601 (1989).
5. Rosen, S., Walliker, J.R., Fourcin, A.J. & Ball, V. *J. Rehab. Res. Dev.* **24**, 239–260 (1988).
6. Moore, B.C.J. in *Cochlear Implants* (ed. Gray, R.F.) 163–179 (Croom Helm, London, 1985).
7. Licklider, J.C.R. *Experientia* **7**, 128–134 (1951).
8. Smoorenburg, G.F. *J. Acoust. Soc. Am.* **48**, 924–941 (1970).
9. Patterson, R.D. *J. Acoust. Soc. Am.* **82**, 1560–1586 (1987).
10. Meddis, R. & Hewitt, M. in *Basic Issues in Hearing* (eds Duifhuis, H., Horst, J.W. & Wit, H.P.) 148–153 (Academic, London, 1988).
11. Howard, I.S. and Walliker, J.R. *Eurospeech, Paris* (in the press).

SEA-level rise is one of the most disturbing possible consequences of global 'greenhouse' warming. Estimates made in the past decade suggest that a rise of about one metre (or possibly more) will occur by the time of doubling of carbon dioxide and other trace gases in the atmosphere. At a recent symposium on this subject*, however, the consensus was that these estimates may be too high. A rise of only about 0.3 m is the best estimate but the uncertainties involved allow the possibility of a rise of as much as 0.7 m or perhaps none at all.

Changes in global average sea level, on timescales of a decade to a century, are caused by three main processes: changes in ice mass on land, changes in the temperature of ocean water, and changes in liquid water stored on land in groundwater aquifers or surface reservoirs. The greenhouse connection arises because the first two processes are sensitive to climate change. During the past 50 years or so, sea level has been rising at an average rate of 2.4 ± 0.9 mm yr⁻¹ (W. R. Peltier, University of Toronto), a value obtained by correcting tide-gauge data for the continuing glacial isostatic adjustment. Other recent estimates range from about 1.3 to 2.0 mm yr⁻¹.

Explaining the current rise, however, is no trivial task. Thermal expansion accounts for about 0.2 ± 0.1 mm yr⁻¹ (K. Bryan, Princeton University; T. M. L. Wigley, University of East Anglia). The volume of liquid water stored in the ground, in lakes and in reservoirs has probably contributed 0.2 mm yr⁻¹ or less (my own work). And changes in ocean-basin volume normally produce rates of sea-level change of 10^{-3} mm yr⁻¹ or less (C. G. A. Harrison, University of Miami). The small glaciers of the world (all glaciers other than the Greenland and Antarctic Ice Sheets) account for 0.5 ± 0.3 mm yr⁻¹ of sea-level rise averaged over the past 60–80 years (M. Kuhn, Universität Innsbruck; see my paper in *Science* **226**, 1418–1421; 1989).

On the other hand, the Greenland Ice Sheet seems to be thickening, according to recent satellite surveys, at a rate equivalent to 0.45 ± 0.25 mm yr⁻¹ of sea-level lowering (J. Zwally, NASA-Goddard Space Flight Center; *Science* **246**, 1589–1591; 1989). However, this figure is not a long-term average, includes extrapolation over more than half of the ice sheet that lies beyond the satellite coverage, and incorporates little information on the marginal zone of high melt rates. Field measurements of input (snow accumulation) minus output (discharge of ice to the

sea) suggest that the Antarctic Ice Sheet is currently growing at a rate equivalent to 0.75 ± 0.25 mm yr⁻¹ of sea-level reduction (C. R. Bentley, University of Wisconsin), but different values have been suggested by other authors. Comparison of these figures with those for the sea-level rise makes it apparent that the cause of the current rise in global sea level is still poorly understood; the error limits for the total change must be understood.

Future changes in the components of sea-level change can be calculated for different models of greenhouse warming, but must be tempered by the uncertainty in our understanding of the present causes. Different authors use somewhat different outlooks for climatic warming, but most are based on a global average temperature rise of 3–5 °C by about the year 2050. Estimates of ocean thermal expansion using a relatively simple energy-balance model (Wigley) or a fully coupled atmosphere–ocean general circulation model (Bryan) are broadly similar and concur with earlier results — 0.2 ± 0.1 m of total rise.

The new calculations of the glacier contribution to sea-level rise, on the other hand, are considerably reduced over earlier estimates. By 2050, small glaciers and ice caps are expected to contribute 0.16 ± 0.14 m, the Greenland Ice Sheet 0.08 ± 0.12 m, and the Antarctic Ice Sheet -0.3 ± 0.2 m. Adding an estimate of the amount of groundwater depletion, equivalent to 0.2 ± 0.3 m of sea-level rise, gives an estimated total rise of 0.34 ± 0.42 m.

The estimated contributions from glaciers and ice sheets are lower than before for three reasons. First, warmer air over Antarctica and Greenland can hold more moisture and produce greater snowfall. Atmospheric general circulation models include this effect, but these models do not simulate reality over the ice sheets well; the recent results are based on simple physics (change in the saturated vapour pressure over ice) and observation (spatial and temporal relations of accumulation and temperature) (Bentley; S. Warren & S. Frankenstein, *Ann. Glaciol.*, in the press). Second, greenhouse warming will produce increased melting on polar glaciers and ice caps, but much of this meltwater will percolate and refreeze in the subfreezing snow, delaying any runoff into the sea. Recent models (W. T. Pfeffer, University of Colorado) indicate that this delay will last decades for arctic ice caps and the Greenland Ice Sheet. Lastly, although greenhouse warming will probably increase the melting at the undersurface of floating ice shelves in

*Sea Level Change, American Geophysical Union, San Francisco, 6 December 1989.

Antarctica, the dynamic response of the land-based ice sheet and ice streams that feed these ice shelves will probably be so slow that there will be little appreciable impact on sea level by the year 2050 (Bentley). Note, however, that recent studies of the Siple Coast of Antarctica have shown that the velocity of individual ice streams have changed greatly over timescales of decades to centuries.

The consensus estimates reported here have large uncertainties, reflecting both our incomplete knowledge of processes and our lack of sufficient observational data. It does appear that a sea-level rise of

1 m by 2050 is unlikely. But even a 30-cm rise will cause social and economic problems in low-lying areas: this modest rise corresponds to a retreat in shoreline of 30 m or more, unless artificial protection is established. There would also be intrusions of saltwater in estuaries and groundwater aquifers, some destruction of coastal wetlands and an increased frequency of damage from storm surges. □

M. F. Meier is in the Institute of Arctic and Alpine Research and the Department of Geological Sciences, University of Colorado, Boulder, Colorado 80309, USA.

involved in the synthesis of homocitrate, which has recently been shown to be a component of the FeMo-cofactor⁷. Mutants lacking the *nifV* gene product synthesize an MoFe-protein that is less efficient at reducing N₂ (that is, a higher proportion of the electron flux through the protein is wasted as H₂)⁸. Thus the reactivity of the Mo and/or Fe atom that binds and reduces protons, dinitrogen or acetylene is modified by the presence of homocitrate in the FeMo-cofactor.

The reduction of acetylene to ethane is a characteristic not only of the mutant molybdenum nitrogenases constructed by Scott *et al.* but also of the recently discovered vanadium-containing nitrogenases and the less well characterized third nitrogenase that contains Fe but apparently no Mo or V (see R. Cammack's News and Views article⁹). Interestingly, the insertion of the isolated FeV-cofactor into apo-Mo nitrogenase produces a hybrid enzyme that does not reduce N₂ but which does reduce acetylene to a mixture of ethylene and ethane¹⁰. Similarly, addition of molybdenum to the growth medium of a strain of *A. vinelandii* that is capable of synthesizing only the third nitrogenase (the structural genes for the Mo and V nitrogenases were deleted) results¹¹ in a nitrogenase that efficiently reduces acetylene to ethane (45 per cent) but will not reduce N₂. Thus insertion of the 'wrong' cofactor results in changed chemistry at the enzyme's active site.

The chemistry of N₂ reduction under

New ligands bound to work

R.N.F. Thorneley

NITROGENASE, the bacterial metallo-enzyme that reduces dinitrogen (N₂) to ammonia, is important not only because of its role in agriculture, but also because it serves as a model system for studies of the way in which transition-metal-cluster cofactors are assembled and inserted into proteins. The substrate specificity and reactivity of a metal ion or cluster in a protein depends not only on which metal ions are used (Mo or V with Fe in nitrogenase), but also on the way it is attached to the peptide-chain ligands. By using site-specific mutagenesis to alter the chemical environment of the substrate binding site in the nitrogenase from *Azotobacter vinelandii*, Scott *et al.*, who describe their experiments on page 188 of this issue¹, have been able to tune the action of this enzyme.

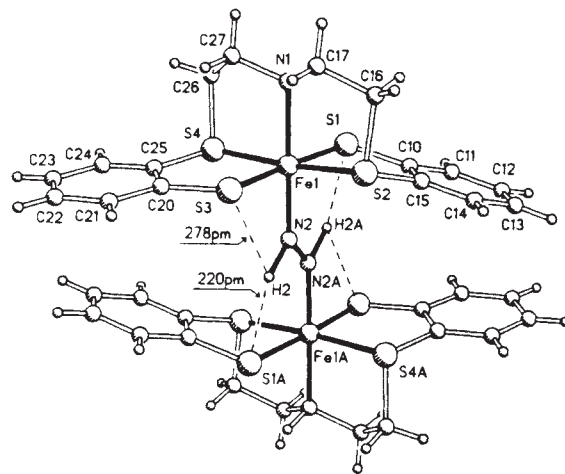
Site-specific mutagenesis of the gene encoding the host protein provides a method of changing or removing a potential ligand. This allows the location of the metal ion or cluster to be determined and the effect of the changed ligation on reactivity to be assessed. This approach is particularly valuable with large multi-meric proteins such as nitrogenases, which contain at least six metal clusters — four 4Fe-4S clusters ('p-centres') and two FeMo-cofactors — and where it has not been possible to isolate the individual subunits with the metal clusters intact. In this way, Scott *et al.* have identified the α -subunits of nitrogenase from *A. vinelandii* ($\alpha_2\beta_2$ structure) as the location of the two FeMo-cofactors that act as substrate reduction sites. They also show that replacement of histidine 195 by an asparagine or of glutamine 191 by a lysine in the α -subunit causes the mutant nitrogenase to reduce acetylene (a substrate analogue of N₂) to a mixture of ethylene and ethane (the wild-type enzyme produces only ethylene). Interestingly, neither of the mutant enzymes reduces N₂, as indicated by the inability of the

mutant strains to grow in the absence of a source of fixed nitrogen.

The authors also recorded the electron-spin-resonance signal that is characteristic of the FeMo-cofactor, finding it to be significantly changed by the introduction of these mutations. Experiments at Sussex University² on the α -subunit of the nitrogenase from *Klebsiella pneumoniae* revealed a similar broadening when cysteine 275 was replaced by alanine. This change also caused the FeMo-cofactor to be bound less tightly to the protein. These data^{1,2} show that histidine 195, glutamine 191 and cysteine 275 are involved in maintaining the structural integrity of the FeMo-cofactor binding site and may be directly bound to the cofactor. Electron-spin-echo studies³ have previously shown that one or more of the ligands to the FeMo-cofactor in the enzyme is a nitrogen base, but the particular residue or residues could not be identified.

In the new work, Scott *et al.*¹ selected the residues to be mutated by comparing the DNA sequences coding for the α - and β -subunits of nitrogenase (*nifD* and *K* genes) with those of two other genes (*nifE* and *N*) which are known to be involved in FeMo-cofactor synthesis⁴. The rationale is that the *nifE* and *N* gene products, which like nitrogenase form a protein with an $\alpha_2\beta_2$ structure⁵, may show sequence homology at an FeMo-cofactor assembly or binding site; residues 180–201 of the nitrogenase α -subunit (*nifD*) are strongly homologous to residues 150–170 of the *nifE* gene product⁶.

The *nifV* gene product is also necessary for the synthesis of fully active FeMo-cofactor. The product of this gene is



Molecular structure of the nitrogenase model compound synthesized by Sellmann *et al.* showing NH...S hydrogen bonds. (From ref. 13.)

ambient conditions by Mo (or W) complexes is well characterized¹². However, the Mo in these complexes is usually attached to non-physiological phosphine ligands and is not part of a cluster containing iron and sulphur (unlike the FeMo-cofactor of nitrogenase). The likelihood that nitrogen and sulphur ligands bind the FeMo-cofactor to the nitrogenase α -subunit and the possible exis-

tence of a nitrogenase containing only Fe, make new work by Sellmann *et al.*¹³ extremely interesting. These authors have synthesized for the first time a compound with an Fe atom in a sulphur and nitrogen ligand environment to stabilize a proposed intermediate in N₂ reduction. A diazene group (=NH=N=), formed by the oxidation of hydrazine, bridges two Fe atoms, each of which is surrounded by four sulphur and one nitrogen ligand. The X-ray structure analysis (see figure) shows that the *trans*-N₂H₂ ligand is stabilized by hydrogen bonding to the sulphur atoms.

After consideration of the calculated energy of the NH...S hydrogen bond formed between the peptide amide groups and cysteinyl sulphurs that ligate the 4Fe-4S cluster in ferredoxins, Sellmann *et al.* suggest that intermediates in N₂ reduction (such as N₂H₂) could be stabilized by hydrogen bonding to sulphur atoms at the active site of nitrogenase (presumably in the FeMo-cofactor). The recognition of this method of stabilizing intermediates in N₂

reduction is probably more significant than the use of the compound as a model for the active site of a nitrogenase, whether it contains Mo, V or only Fe. □

R.N.F. Thorneley is at the AFRC Institute of Plant Science Research, Nitrogen Fixation Laboratory, University of Sussex, Brighton, BN1 9RQ, UK.

1. Scott, D.J., May, H.D., Newton, W.E., Brigle, K.E., & Dean, D.R. *Nature* **343**, 188–190 (1990).
2. Kent, H.M., Ioannidis, I., Gormal, C., Smith B.E. & Buck, M. *Biochem. J.* **264**, 257–264 (1989).
3. Thomann, H. *et al. J. Am. chem. Soc.* **109**, 7913–7914 (1987).
4. Roberts, G.P., MacNeil, T., MacNeil, D. & Brill, W.J. *J. Bact.* **136**, 267–279 (1978).
5. Paustian, T.D., Shah, V.K. & Roberts, G.P. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6082–6086 (1989).
6. Brigle, K.E., Weiss, M.C., Newton, W.E. & Dean, D.R. *J. Bact.* **169**, 1547–1553 (1987).
7. Hoover, T.R., Imperial, J., Ludden, P.W. & Shah, V.K. *Biochemistry* **28**, 2768–2771 (1989).
8. McLean, P.A. & Dixon, R.A. *Nature* **292**, 655–656 (1982).
9. Cammack, R. *Nature* **333**, 595–596 (1988).
10. Smith, B.E., Eady, R.R., Lowe, D.J. & Gormal, C. *Biochem. J.* **250**, 299–302 (1988).
11. Pau, R.N., Mitchenall, L.A. & Robson, R.L. *J. Bact.* **171**, 124–129 (1989).
12. Henderson, R.A., Leigh, G.J. & Pickett, C.J. *Adv. inorg. Chem. Radiochem.* **27**, 197–292 (1983).
13. Sellmann, D., Soglowek, W., Knock, F. & Moll, M. *Angew. Chem. Int. Ed. Engl.* **28**, 1271–1272 (1989).

QUASAR ASTRONOMY

The Universe at high redshift

Bob Carswell and Paul Hewett

In two papers^{1,2} in the *Astronomical Journal* Schneider, Schmidt and Gunn present optical spectra of the quasar Q1158+4635, redshift $z = 4.73$ and ten further quasars, redshifts $z > 4$, selected from their own and other recent surveys. The discovery of Q1158+4635, the new high-redshift record holder, confirms the ability of astronomers to undertake highly effective searches for quasars at these extreme redshifts. The two largest systematic surveys of high-redshift quasars^{3,4}, now comprise more than 100 objects with redshifts $z > 3$. Completion of the analyses of these samples will place significant constraints on the change in the space density of the quasar population at high redshift, and this provides important information on the epoch of collapse of massive bound systems.

Definite conclusions regarding the evolution of the quasar population await the results of these detailed and involved statistical analyses. However, the quasar spectra presented by Schneider *et al.*, may provide dramatic new information on the physical conditions of the intergalactic medium at very early times. Radiation from a quasar, or any background source, is modified by material along our line of sight and quasar spectra contain an impressive variety of absorption systems resulting from the presence of intervening concentrations of mass. The unambiguous identification of the objects responsible for most of the absorption systems has

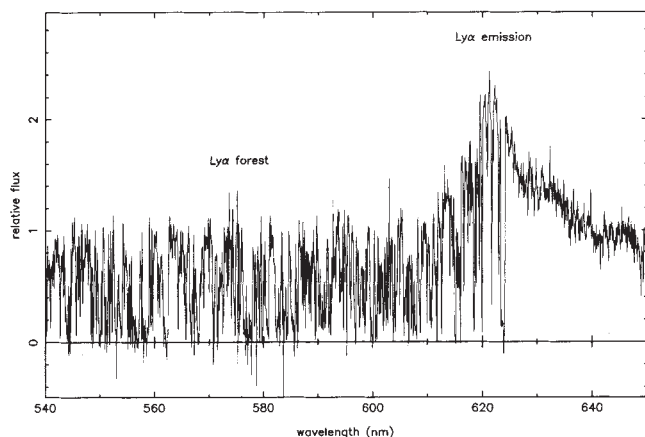
proved elusive. The strong resonance-transition absorption features, both from hydrogen (the 121.6-nm Lyman- α transition) and from the common heavier elements (the 154.8–155.0-nm C³⁺ doublet for example), occur in the rest-frame ultraviolet and can be observed from the ground at redshifts only significantly in excess of 1 (so that their observed wavelengths are at least doubled). This lower redshift limit is set by the strong ultraviolet absorption occurring in the Earth's atmosphere. By contrast our ability to identify even intrinsically luminous galaxies is limited to redshifts significantly below unity. Only in the case of magne-

sium (employing the Mg⁺ doublet at 279.6–280.3 nm, which can be observed at low-redshift in quasar spectra) has it been possible to undertake imaging and spectroscopic observations for any optical counterparts to particular absorption systems.

By far the most prevalent absorption systems are the Lyman- α features which arise from the absorption of photons from the quasar by systems containing neutral hydrogen at lower redshift. The individual features are weak and their velocity widths are small but the cumulative effect of the hundreds of features present in any high-redshift quasar spectrum is dramatic; high-resolution quasar spectra at wavelengths below the 121.6-nm Lyman- α emission feature are so dominated by the absorption that the region is termed the Lyman- α forest. At low resolution the individual systems combine to produce an apparent drop in the quasar continuum level to the blue of Lyman- α emission feature (see figure). Analyses of high-resolution quasar spectra have resulted in the determination of the overall properties of the Lyman- α cloud population, including the redshift dependence of the numbers along a sight-line. Observations covering redshifts $2 < z < 3.4$ are well reproduced by a population of Lyman- α clouds with similar physical properties but whose number-density as a function of redshift increases rapidly with redshift.

Theoretical models encompass several mechanisms for the origin and confinement of the clouds but the most promising ones invoke some level of pressure confinement from a hot, ionized intergalactic medium (IGM), the existence of which is supported by observation. Any significant column density of neutral hydrogen in this IGM along our line of sight would result in a general depression of the quasar continuum — the long-sought Gunn–Peterson effect. However, at the highest resolution, spectral regions free of Lyman- α clouds appear and the quasar spectra attain the predicted, absorption-free,

A high-resolution spectrum⁵ of a high-redshift ($z = 4.11$) quasar showing the forest of Lyman- α absorption lines which give rise to an apparent depression of the spectrum when observed at low resolution.



continuum levels, so either the general IGM is very highly ionized or there is very little of it.

From their new low-resolution spectra, Schneider *et al.* have measured the continuum depression for ten quasars and contrast the results with previous continuum-depression measures for redshifts $z < 3.4$. They find that the size of the depression significantly exceeds that predicted by a simple extrapolation of the increase in the number of Lyman- α clouds as a function of redshift based on low redshift observations and propose three possible explanations. First, the number of Lyman- α clouds increases even more rapidly with redshift for redshifts $z > 3.5$. Second, the intrinsic properties of the clouds change at high redshifts so that each produces more absorption. Or third, the excess absorption results not from the Lyman- α clouds, but is due to a new distributed component of neutral hydrogen — the Gunn–Peterson effect.

Schneider *et al.* are understandably confident about the reality of the observations of excess depression, but the measurement is not an easy one to make. The continuum-depression measures at any redshift show a large scatter owing to variations in the number of Lyman- α clouds, the strength of the absorption from each cloud, and the number and strength of lines in the few heavy-element systems present. An additional complication occasionally arises when gas outflow from the quasar gives rise to broad absorption features which have nothing to do with the intervening hydrogen clouds. Furthermore, systematic differences in the range of quasar rest-frame wavelengths used to derive the continuum as a function of redshift and difficulties in achieving good relative fluxes over extended wavelength ranges can be expected to produce systematic differences in the depression measures of the order of 0.1. There is some evidence that depression measures for spectra of quasars at similar redshifts, but obtained and analysed by different groups, show systematic differences of this order.

Thus, although the results presented by Schneider *et al.* are strongly indicative of a change in behaviour at the highest observed redshifts, further detailed studies are required to confirm this. If the explanation of their results is that the continuous absorption from the IGM has been detected, then this means that the degree of ionization of the IGM changes with redshift. Because the ionization is believed to be a result of the radiation flux from a large population (as yet, unseen) of galaxies undergoing an early burst of star formation, the detection of a significant neutral component to the IGM could mean we have seen back to an epoch where most galaxies had not yet formed.

Whatever the impact of the new study on our knowledge of the IGM at redshift $z \approx 4$, the spectra contribute in at least one other area. On the basis of what was thought to be a rapid decline in the space density of quasars at redshifts $z < 3.5$, it has been suggested that increasing absorption by dust at high redshifts results in an apparent decline in the observed space density of quasars. Although the strong selection effect implicit in such a model — that quasars suffering the least extinction are preferentially included in flux-limited samples — has been emphasized, the new study contains quasars selected using several techniques and spanning a factor of around 40 in luminosity. The lack of any correlation between the absolute magnitudes of the quasars and the steepness of their continua (which is a strong predic-

CELL MOTILITY

Multiple microtubule motors

Jonathan M. Scholey

MANY motile activities in eukaryotic cells depend on the action of cytoskeleton-based molecular motors (see table). These motors are enzymes that couple the hydrolysis of ATP to the generation of force and motion relative to cytoskeletal filaments, either actin or microtubules. Until recently, two types of cytoplasmic microtubule-based motors were known. But in a paper in *Cell*¹, Shpetner and Vallee describe a third, which they call 'dynammin'.

Actin-based motors fall into two classes, nonfilamentous myosin-1, which is thought to move membranes along actin filaments and drive active sliding between the filaments, and myosin-2, which forms filaments that interact with actin filaments to cap cell-surface molecules, generate cortical tension, and drive cytokinesis and muscle contraction².

Work during the past few years has revealed that microtubules, which serve as tracks for intracellular particle transport in eukaryotes, interact with two classes of motors that can drive the movement of objects in opposite directions along their surface³ (see figure, *a*). Kinesins transport objects towards the 'plus' (growing) ends of microtubules *in vitro*, whereas cytoplasmic dyneins generally catalyse minus-end-directed transport, and consequently the concerted action of these two motors is thought to mediate bidirectional particle transport within cells⁴.

In addition to supporting the movement of objects along their surface, microtubules also slide along one another both in the interphase cytoplasm⁴ and in the zone of microtubule overlap within the mitotic spindle⁵. Unlike axonemal dyneins⁶, which mediate the sliding of flagellar microtubules, kinesins and cyto-

plasmic dyneins do not seem to be very effective in mediating lengthwise microtubule–microtubule sliding. The new protein described by Shpetner and Vallee¹ can drive active sliding between microtubules *in vitro* (see figure, *b*) and it may, therefore, drive movements within cells that depend on such microtubule–microtubule sliding.

The identification and isolation of dynammin is based on three properties diagnostic of microtubule motors. First, dynammin contains site(s) that can bind reversibly to microtubules in a nucleotide-sensitive fashion; second, it displays a microtubule-activated MgATPase activity; and third, it couples MgATP hydrolysis to microtubule movement as observed in a video-microscope motility assay. Dynammin from calf brain mainly consists of a presumptive catalytic polypeptide of relative molecular mass 100,000 (100k), which has been purified, together with a partially purified cofactor. The 100k polypeptide forms regularly spaced crossbridges between microtubules, causing the polymers to pack into striking hexagonal bundles *in vitro*. Interestingly, when monitored by video-enhanced light microscopy, in the presence of MgATP and the cofactor, these crosslinked microtubule bundles are observed to elongate at rates of $0.2 \mu\text{m s}^{-1}$ or more. Each dynammin crossbridge that drives this elongation is thought to form a structural bond with one microtubule and a mechanochemical bond with an adjacent microtubule within a bundle. Each crossbridge apparently uses energy liberated from ATP hydrolysis to move the mechanochemically bonded microtubule, sliding it along the microtubule tethered by structural bonds at its base.

1. Schneider, D. P., Schmidt, M. & Gunn, J. E. *Astr. J.* **98**, 1507–1522 (1989).
2. Schneider, D. P., Schmidt, M. & Gunn, J. E. *Astr. J.* **98**, 1951–1958 (1989).
3. Warren, S. J., Hewett, P. C. & Osmer, P. S. *Proc. Workshop Optical Surveys for Quasars* (eds Osmer, P. S., Porter, A. C., Green, R. F. & Foltz, C. B.) *Astr. Soc. Pacific Conf. Series* **2**, 97–105 (1988).
4. Schmidt, M., Schneider, D. P. & Gunn, J. E. *Proc. Workshop Optical Surveys for Quasars* (eds Osmer, P. S., Porter, A. C., Green, R. F. & Foltz, C. B.) *Astr. Soc. Pacific Conf. Series* **2**, 87–93 (1988).
5. Webb, J. K. *et al.* *ESO Messenger* No. 51, 15–18 (1988).

tion of the dust obscuration model) means that dust is unlikely to be a significant factor in the decline in the numbers of quasars at the highest redshifts, a conclusion in agreement with recent works on the incidence of dust in the highest column density Lyman- α systems. □

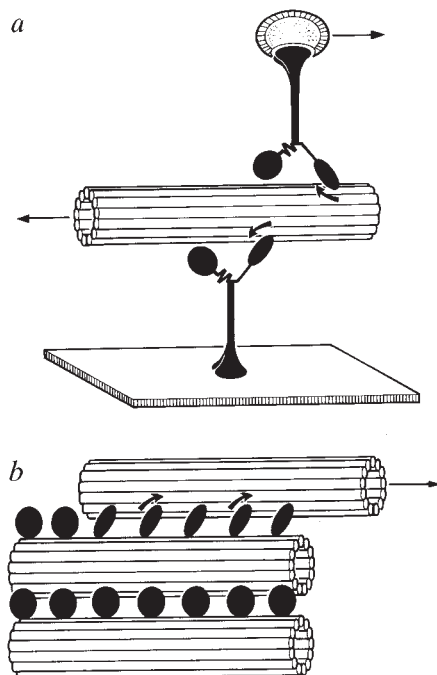
Bob Carswell and Paul Hewett are at the Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK.

An unanswered question concerns the

polarity of this microtubule-microtubule sliding: does dynamin drive the sliding apart of parallel microtubules, like those in axons, or antiparallel microtubules, like those in the mitotic-spindle overlap zone? Another question concerns the identity of the cofactor: preliminary evidence implicates a 70k polypeptide for this role, but the molecular mechanism by which this presumptive regulatory protein turns on the mechanochemical activity of dynamin is unknown.

Dynamin differs from kinesin and cytoplasmic dynein in its polypeptide composition, nucleotide sensitivity of microtubule binding, generation of microtubule sliding and requirement for the cofactor, suggesting that it truly is a novel mechanoenzyme and not, perish the thought, a proteolytic fragment of kinesin or dynein, although future sequence studies may reveal interesting similarities to other microtubule motors.

What are the biological functions of dynamin, kinesin and cytoplasmic dynein? There is growing evidence from motility assays (reviewed in ref. 6), immunocytochemistry^{7,8} and antibody-disruption experiments^{9,10} that kinesin and dynein participate in the movement of membranous tubules and vesicles along microtubules within the cytoplasm at interphase. A role for kinesin in organelle



Cartoon showing the types of movement that are driven by cytoplasmic microtubule motors. *a*, Kinesin and cytoplasmic dynein (not shown) transport particles along microtubules and move microtubules over glass surfaces. *b*, Dynamin forms regular crossbridges that drive active sliding between microtubules (arrow), as described in the text. The drawings are schematic only. The mechanisms of these movements are not known. For details of the organization of dynamin crossbridges see ref. 1.

Classification of molecular motors from eukaryotic cells

Actin-based		
	Description	Function
Myosin-1	Monomeric ~110–140k heavy chain	Moves membranes on actin filaments. Drives active sliding between actin filaments.
Myosin-2	Filaments ~170–240k heavy chain	Muscle contraction, cortical tension, capping surface molecules, cell polarity, cytokinesis.
Microtubule-based		
	Description	Function
Axonemal dynein	~400–500k heavy chain	Drives active sliding between flagellar microtubules. Moves glass surfaces towards 'minus' end of microtubule.
Cytoplasmic dynein	~400–500k heavy chain	Moves glass surfaces, anionic beads and membranes towards 'minus' end of microtubule.
Kinesin	110–140k heavy chain	Moves glass surfaces, anionic beads and membranes towards 'plus' end of microtubule.
Dynamin	100k heavy chain + cofactor	Drives active sliding between microtubules.

transport is supported by the observation that *Drosophila* containing mutant kinesin display phenotypes that are consistent with neuromuscular defects resulting from impaired anterograde transport of axoplasmic organelles¹¹. Whether dynamin contributes to the active sliding between microtubules seen in the cytoplasm of some interphase cells⁴ is an interesting question that merits attention.

In dividing eukaryotic cells, chromosome movement is coordinated by microtubules of the mitotic spindle, and may involve the action of molecular motors. There has been only limited success in attempts to identify these motors. Kinesin, for example, is strikingly associated with the mitotic spindle in some cells¹² but has not been detected in the mitotic apparatus of others^{7,8}. These different results could reflect variations in the kinesin-bound membrane content of different spindles or the presence of a widespread and functionally important spindle motor whose abundance or accessibility to antibody precluded detection in the latter studies^{7,8}. But, as direct functional evidence is not yet available, the significance of kinesin in the spindle is unknown.

Shpetner and Vallee¹ focus attention on dynamin as another candidate for a mitotic motor. In particular, depending upon the relative polarity of the microtubules that it slides apart, dynamin could be an attractive candidate for a motor that drives the poles of the mitotic spindle apart during anaphase¹. The presence of a dynamin-like enzyme in sea-urchin eggs¹³, where mitosis is a primary function of microtubules, is consistent with such a role in mitosis for dynamin. However, as with the other motors, direct evidence of the function of dynamin in dividing cells is needed.

For some microtubule motors, such

functional information may already be emerging. Exciting recent results strongly suggest that microtubule motors are required for some cells to complete their cell division or reproduction cycles. Rose and co-workers¹⁴, for example, find that the derived amino-acid sequence of the yeast *KAR3* gene predicts that its protein product is related to kinesin¹⁵, and they propose that it serves as a motor for microtubule-driven nuclear fusion and perhaps mitosis as well. Morris and collaborators (UMDNJ/Robert Wood Johnson Medical School) have discovered an *Aspergillus* gene encoding a kinesin-like motor that is involved in spindle pole body separation and so must be required for nuclear division (R. Morris, personal communication).

How many different microtubule motors do cells contain? Currently three classes of cytoplasmic microtubule motors have been purified and characterized — kinesins, cytoplasmic dyneins and dynamin — but there may well be more. It will now be interesting, for example, to determine whether *KAR3* and the *Aspergillus* motor are novel mechanochemical proteins, true kinesins or more closely related to other motors such as dynamin. To determine the structural and functional relationships between different motor molecules requires further cloning, sequencing and characterization of the genes that encode them, together with a detailed analysis of the properties of the corresponding protein products. Recent studies reveal the presence in some cells of multiple myosins with different functions (obtained by fusing a motor domain to different tails)². Therefore it would not be surprising if a sizeable family of microtubule motors were to be uncovered, each involved in specific motile events. ►

What are the mechanisms of microtubule movements? The question of how molecular motors couple ATP hydrolysis to the production of force has traditionally been addressed using biochemical, kinetic and structural studies¹⁶, but many workers are now adding video-enhanced light-microscope motility assays to their armoury of techniques for analysing the mechanism of mechanochemical coupling. Thus, microscopic methods have now been used to characterize motor enzymology in detail¹⁷, to measure the sizes of displacements that may correspond to the steps that motors take as they track along their filament¹⁸, and in an exciting development by Howard *et al.*¹⁹ published in a recent issue of *Nature*, to make visible the activity of single motor molecules.

Investigators are also starting to apply optical, biochemical and structural assays to wild-type and mutant motor molecules and their fragments, expressed in transformed bacteria or *in vitro* from cloned genes^{15,20}. This combination of methods should ultimately provide a detailed picture of the molecular events involved in mechanochemical coupling, and may reveal a basic force-generating mechanism that is common to each type of microtubule motor. Then it would be fascinating to know if multiple microtubule motors result from a form of combinatorial joining of conserved motor domains to an assortment of tails, with each type of tail binding a subset of the diverse structures that are moved relative to microtubules in cells. □

Jonathan M. Scholey is in the Department of Zoology, University of California, Davis, California 95616, USA.

When seven's a crowd

Charles Bailyn and Michael Garcia

THE X-ray sky is a variable place. Unlike the optical heavens, where large irregular intensity variations are confined to very rare or very faint objects, the dozens of X-ray transients that have been observed in the 25-year history of X-ray astronomy appear to spring up from non-detectable limits to be among the brightest objects in the sky. On page 148 of this issue¹, Koyama *et al.* report on the discovery of seven new hard X-ray transients, which are unexpectedly concentrated in a region just a few square degrees in area.

Because X-rays do not penetrate the atmosphere, X-ray astronomy must be conducted from space, primarily by orbiting observatories. Transient sources, whose locations cannot be predicted, are best detected by satellites that periodically reobserve likely spots in the sky, or that continuously scan the heavens with 'all-sky monitors' (ASMs). The recent discoveries were made with the large-area counters on the Japanese orbiting X-ray observatory Ginga² (Fig. 1), using the former technique. Ginga has a remarkable record of detecting transients, partly because of the instrument's large field of view and high sensitivity (especially to relatively high-energy X-rays), but also because of the Ginga team's commitment to performing the repeated scans of the same region required to pick up transients.

The so-called hard X-ray transients, which are distinguished by the relatively high-energy X-rays they emit, are thought to be binary stars comprising a neutron star and a massive blue main-sequence star³. The neutron star accretes part of the

Dramatic changes in the accretion rates are required to explain the transient nature of many X-ray pulsars — exactly what drives these changes remains mysterious.

Before the new Ginga discoveries it seemed reasonable to assume that these pulsars were uniformly distributed in the Galaxy⁴. However, the seven new transients are strongly concentrated at a galactic longitude of 30° (Fig. 2). Koyama *et al.* note that this is the direction of the '5-kiloparsec' spiral arm of our Galaxy and

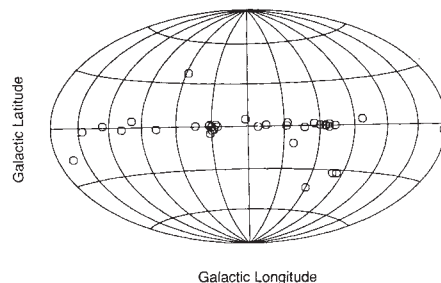


FIG. 2 The distribution of transient X-ray pulsars in galactic coordinates shows the newly found concentration of seven pulsars at $l \approx 30^\circ$ (immediately left of centre).

suggest that these newly found transients are embedded in that arm. Other evidence points to the 5-kiloparsec arm as the site of a burst of star formation within the past 10 million years⁵. The new discoveries support this idea, because the bright blue companions of the neutron stars in transients would have burnt themselves out if they were much older than this. Interestingly, recent analysis of the Einstein observatory's X-ray observations of nearby spiral galaxies shows concentrations of X-ray sources near clouds of ionized hydrogen⁶, also probably the site of recent star formation.

Because of the small number of X-ray ASMs flown and the low frequency of Ginga scans of the galactic plane, the duty cycle of transient X-ray pulsars remains largely unknown. Therefore large and poorly known correction factors are necessary when estimating the total number of transients in the Galaxy. The discovery that the transients are concentrated in star-forming regions and Ginga's high success rate in finding new ones with each additional scan indicate that the total number of transients in the Galaxy may be much higher than previous estimates of about a thousand⁷. However, further observations to constrain recurrence times and duty cycles will be required before accurate estimates of the numbers of these objects can be made.

The ASMs soon to fly (aboard the US XTE and the Soviet Spectrum-X Gamma X-ray astronomy satellites) will be hard

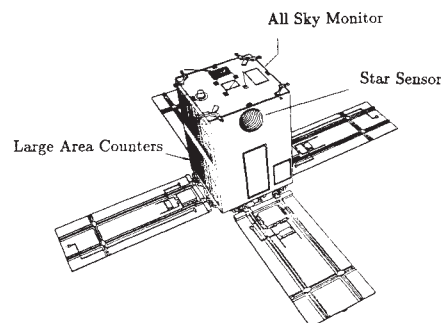


FIG. 1 Schematic diagram of the Japanese X-ray astronomy satellite Ginga. (From ref. 8.)

stellar wind blown off by the blue star, heating the accreted gas sufficiently to create the X-ray emission. The rotation of the highly magnetized neutron star can often be seen in the form of X-ray pulses. The observed changes in rotation rate, which can be either positive or negative, are presumably influenced by the angular momentum of the accreted material.

1. Shpetner, H.S. & Vallee, R.B. *Cell* **59**, 421–432 (1989).
2. Spudich, J.A. *Cell Regulation* **1**, 1–11 (1989).
3. McIntosh, J.R. & Porter, M.E. *J. biol. Chem.* **264**, 6001–6004 (1989).
4. Koonce, M.P., Tong, J., Euteneur, U. & Schliwa, M.P. *Nature* **328**, 737–739 (1987).
5. Cande, W.Z. & McDonald, K.L. *Nature* **316**, 168–170 (1985).
6. Schroer, T.A., Steuer, E.R. & Sheetz, M.P. *Cell* **56**, 937–946 (1989).
7. Pfister, K.K., Wagner, M.C., Stenoien, D.L., Brady, S.T. & Bloom, G.S. *J. Cell Biol.* **108**, 1453–1463 (1989).
8. Hollenbeck, P.J. *J. Cell Biol.* **108**, 2335–2342 (1989).
9. Brady, S.T., Pfister, K.K. & Bloom, G.S. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
10. Hollenbeck, P.J. & Swanson, J.A. *J. Cell Biol.* **109**, 81a (1989).
11. Saxton, W.M. & Raff, E.C. *J. Cell Biol.* **109**, 281a (1989).
12. Scholey, J.M., Porter, M.E., Grissom, P.M. & McIntosh, J.R. *Nature* **318**, 483–486 (1985).
13. Collins, C.A. & Vallee, R.B. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4799–4803 (1986).
14. Vogel, J.P., Meluh, P.B. & Rose, M.D. *J. Cell Biol.* **109**, 192a (1989).
15. Yang, J.T., Laymon, R.A. & Goldstein, L.S.B. *Cell* **56**, 879–889 (1989).
16. Johnson, K.A. *Rev. Biophys. Chem.* **14**, 161–168 (1985).
17. Cohn, S.A., Ingold, A.L. & Scholey, J.M. *J. biol. Chem.* **264**, 4290–4297 (1989).
18. Gelles, J., Schnapp, B.S. & Sheetz, M.P. *Nature* **331**, 450–453 (1988).
19. Howard, J., Hudspeth, A.J. & Vale, R.D. *Nature* **342**, 154–158 (1989).
20. de Cuevas, M., Yang, J.T. & Goldstein, L.S.B. *J. Cell Biol.* **109**, 281a (1989).

pressed to detect transient pulsars at the low flux levels found with the Ginga scans of the galactic plane. There are a number of focusing X-ray telescopes slated for launch in the 1990s including the West German-US-UK satellite ROSAT, the

1. Koyama, K. *et al.* *Nature* **343**, 148–149 (1990).
2. Special Issue for Ginga *Pub. astr. Soc. Jap.* **41**, 345–426 (1989).
3. Maraschi, L., Treves, A. & van den Heuvel, E. P. J. *Nature* **259**, 292 (1976).
4. Nagase, F. *Pub. astr. Soc. Jap.* **41**, 1–79 (1989).
5. Hayakawa, S., Ito, K., Matsumoto, T. & Uyama, K. *Astr. Astrophys.* **58**, 325–330 (1977).
6. Trinchieri, G. in *The Origin and Evolution of Neutron Stars* (eds Helfand, D.J. & Huang, J.H.) 149–159 (Reidel, Dordrecht, 1987).
7. Rappaport, S. & van den Heuvel, E.P.J. in *Be Stars, IAU Symp.* 98 (eds Jaschek, M. & Groth, H.) 327–344 (Reidel, Dordrecht, 1982).
8. Turner, M.J.L. *et al.* *Pub. astr. Soc. Jap.* **41**, 347 (1989).

Japanese satellite ASTRO-D, the US AXAF and European XMM missions which will have the sensitivity to detect these objects. But because these have relatively narrow fields of view, they cannot efficiently scan the galactic plane searching for new transients. Now that we know that transient X-ray pulsars may be concentrated in small star-forming regions, however, these telescopes could monitor these parts of the sky and help to determine the duty cycles and absolute number of these enigmatic objects. □

Charles Bailyn and Michael Garcia are at the Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.

MOLECULAR GENETICS

Mice, men and sickle cells

D.J. Weatherall

WHEN trying to press the virtues of molecular medicine to sceptical clinicians or granting bodies, the story of sickle cell anaemia usually comes low on the agenda. For although the molecular basis of this often distressing condition was established by Vernal Ingram over 30 years ago, we still can do little more for our patients than could William Herrick when he first identified the disorder in a West Indian student in Chicago in 1910. One of the main constraints on research on this complex, multisystem disease has been lack of a good animal model. So workers in the field will be excited to read the title of the paper by David Greaves and his colleagues on page 183 of this issue of *Nature*¹, which states that we now have a mouse model of sickle cell anaemia.

The facts of the disease are deceptively simple. Defective haemoglobin (S) differs from normal haemoglobin (A) by the substitution of valine for glutamic acid at position 6 in the β -globin chain². In concentrated haemoglobin solutions that are partially or fully deoxygenated this change leads to polymerization of haemoglobin molecules and the formation of intracellular fibres which cause the sickle cell deformity. The deformity results in reduced flexibility of the red cell and hence in its impaired passage through the microcirculation³. This rheological abnormality has two results. First, red cells do not survive as long as they do normally, resulting in chronic haemolytic anaemia. Second — and more importantly — small blood vessels become blocked by aggregates of sickled erythrocytes, leading to tissue damage.

If this were all, sickle cell anaemia would be a relatively mild disorder; the chronic anaemia is tolerated quite well and tissue damage may be minimal. But

unfortunately the disease is often complicated by episodes called 'crises'. These acute exacerbations take several forms, including widespread and agonizing bone pain from infarction of the marrow, blockage of larger blood vessels to the lung or brain with sickled erythrocytes, massive sequestration of sickle cells in the spleen or liver leading to profound anaemia, or equally severe anaemia due to bone marrow failure following virus infection.

Patients with sickle cell anaemia are also prone to severe infection, the commonest cause of death at all ages⁴. The cause of the painful and sequestration crises is unknown, although they probably reflect complex interactions between sickled erythrocytes and the vascular endothelium together with variability of response of the microcirculation in different organs to environmental changes such as body-surface temperature. Similarly, except for defective spleen function which is a result of repeated infarction ('autosplenectomy'), the reason for the increased susceptibility to infection is not understood.

Another problem is the variable clinical course of the disease. Some patients suffer repeated crises and severe organ damage, whereas others go through life with little disability. Although a few genetic modifiers of the sickle cell phenotype have been identified^{3,4}, the reasons for such clinical heterogeneity are not understood. This is of particular relevance now that the disease can be identified early in fetal life; counselling parents at risk for having children with the condition is particularly difficult.

The apparent *tour de force* reported by Greaves *et al.*¹ resulted from the recent identification of the dominant control region (DCR) sequences which flank the

human β -globin locus and which direct high-level, copy-number-dependent expression of the human β -globin gene in erythroid cells in transgenic mice⁵. By inserting a construct which included two human α genes and the defective human β^s -globin gene, all driven by the DCR, it was possible to produce two mice that have relatively high levels of human haemoglobin S in their red cells. *In vitro*, these cells showed typical sickling in conditions of reduced oxygen tension; and the mice appear to have occasional irreversibly sickled cells in their circulation, showing that there may be some intravascular sickling *in vivo*.

These observations confirm that it is possible to obtain high levels of expression of human globin genes in mouse erythrocytes using the β -globin DCR¹. But although the results are of considerable interest, a great deal more work is needed to find out how closely the mouse model resembles human sickle cell anaemia. In the two mice that have relatively large amounts of sickle cell haemoglobin in their red cells, the haematological findings at 4 and 15 days are indistinguishable from those of normal animals. In particular the transgenic mice are not anaemic and the reticulocyte count, a relatively sensitive indicator of red-cell turnover, is not elevated, indicating that their red cells are surviving as normal. It may be necessary to breed mice with higher levels of haemoglobin S in their red cells and to determine whether this leads to anaemia, tissue damage and the other protean manifestations of human sickle cell anaemia. It may turn out, of course, that the size of the red cells, the circulation dynamics and the regulation of the microcirculation are so different in mice and men that this will not be the case. Even so, it may still be possible to use the mouse model for testing anti-sickling agents.

Those trying to understand the pathology of sickling disorders will wait for the results of more prolonged observation of these transgenic animals with great interest. But it is far too early to say whether we have an animal model of sickle cell anaemia. This research field has had more than its fair share of false leads and disappointments — it is to be hoped that this will not be yet another. □

D.J. Weatherall is in the Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK.

1. Greaves, D.R. *et al.* *Nature* **343**, 183–185 (1990).
2. Ingram, V.M. *Nature* **178**, 792 (1956).
3. Schechter, A.N., Noguchi, C.T. & Rodgers, G.P. in *The Molecular Basis of Blood Diseases* (eds Stamatoyannopoulos, G., Nienhuis, A.W., Leder, P. & Majerus, P.W.) 179–218 (Saunders, Philadelphia, 1987).
4. Serjeant, G.R. *Sickle Cell Anaemia* (Oxford University Press, 1985).
5. Grosfeld, F., Blom van Assendelft, G., Greaves, D.R. & Kollias, G. *Cell* **51**, 975–985 (1987).

Telomeres sans frontières

Elizabeth H. Blackburn

THE ciliated protozoa have sometimes been suspected of indulging in strange practices unique to themselves. A paper in *Cell* by Gregg Morin¹ shows that, when it comes to telomere synthesis, this suspicion is unfounded. Morin reports the identification of telomere terminal transferase (telomerase) activity in crude HeLa cell extracts. Telomerase is a ribonucleoprotein enzyme that synthesizes telomeric DNA²⁻⁶. The finding of activity in human cells similar to the telomerases hitherto known only in ciliates highlights the extraordinary conservation of telomere structure and function among eukaryotes.

Telomeres, the specialized DNA-protein complexes at the ends of the eukaryotic chromosome, are required for chromosome stability. Evidence has been accumulating that the molecular features of telomeres are universal. Telomeric DNA structure was first studied in the 'lower' eukaryotes but now it is known to be conserved between widely divergent lower and higher eukaryotes^{7,8}, including humans⁹. Telomeric DNA consists of simple, tandemly repeated sequences, usually G+C-rich, characterized by clusters of G residues and an overall strand-composition asymmetry resulting in G-rich and C-rich strands, with the G-rich strand always being orientated 5' to 3' towards the chromosome end and protruding as a 3' overhang. The sequence AGGGTT has emerged as the most popular eukaryotic telomeric repeat, cropping up in corners as far flung as trypanosomes, acellular slime moulds, filamentous fungi and humans.

Ciliated protozoa are a rich source of telomeres and have been particularly useful for studies of telomere structure and metabolism. Telomerase, first discovered in the ciliate *Tetrahymena*²⁻⁴, is a ribonucleoprotein enzyme with essential RNA and protein components^{3,4}. The *Tetrahymena* enzyme adds tandem repeats of the telomeric sequence of this species, TTGGGG, onto the 3' end of a telomeric oligonucleotide primer²⁻⁴. Such addition occurs independently of any added nucleic acid template². Similar results subsequently emerged from study of the telomerases of the ciliates *Oxytricha*⁵ and *Euplotes*⁶, which synthesize analogous species-specific telomeric repeats. Now, Morin has extended these observations to include telomerase activity in human cell extracts, which synthesizes AGGGTT repeats. The human activity is also sensitive to ribonuclease, suggesting that, like the ciliate telomerases, it too is a ribonucleoprotein. Its primer recognition properties are also similar to those of the ciliate telomerases. The telomerase RNA moieties of

Tetrahymena and *Euplotes* each contain a telomere-complementary sequence that acts as the template for addition of species-specific repeats (ref. 4 and our unpublished results). Given the evolutionary conservation of telomerase activity shown by Morin's work, we can expect that the RNA activity of the human enzyme will have an analogous role.

Human cells have also been the source of other interesting findings about telomeres. De Lange *et al.*¹⁰ report that, in chromosome-specific sequences next to the terminal AGGGTT repeats, DNA modification occurs in a developmentally controlled manner, with telomeric-region DNA from sperm being undermodified compared with these regions in somatic cells. Modification of some trypanosome telomeric regions has also been observed to be developmental-stage-specific, as well as gene-expression-specific, although the relationship between the modifications in these two systems is not clear. The work of De Lange *et al.* also indicates that Morin's choice of cells may have been a fortunate one: telomeres in one HeLa cell line are exceptionally long¹⁰, which may suggest that, in this HeLa cell line at least, telomerase activity is relatively high.

Telomeres can freely cross the borders of eukaryotic phyla and be used, *in vivo*, in very distantly related species. This behaviour contrasts dramatically with the species specificity of other chromosomal elements such as centromeres and replication origins. It was shown last year that human telomeres function in yeast (see ref. 1), again supporting the notion that shared properties are not restricted to the telomeres of lower eukaryotes. This passportless behaviour of telomeres *in vivo* may have as its basis the common primer recognition properties of telomerases, because all the telomerases studied can recognize as a primer every G-rich strand telomeric sequence that has been tested^{1,3,5,6}. Morin's work further reminds us that, in the world of telomeric DNA, there are no Berlin walls between species. □

Elizabeth H. Blackburn is in the Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA.

1. Morin, G. *Cell* **59**, 521-529 (1989).
2. Greider, C.W. & Blackburn, E.H. *Cell* **43**, 405 (1985).
3. Greider, C.W. & Blackburn, E.H. *Cell* **51**, 887 (1987).
4. Greider, C.W. & Blackburn, E.H. *Nature* **337**, 331-337 (1989).
5. Zahler, A.M. & Prescott, D.M. *Nucleic Acids Res.* **16**, 6953-6972 (1988).
6. Shippen-Lentz, D. & Blackburn, E.H. *Molec. Cell. Biol.* **9**, 2761-2764 (1989).
7. Blackburn, E.H. & Szostak, J.W. *A. Rev. Biochem.* **53**, 163-194 (1984).
8. Richards, E.J. & Ausubel, F.M. *Cell* **53**, 127-136 (1988).
9. Moyzis, R.K. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 6622-6626 (1988).
10. de Lange, T. *et al. Molec. Cell. Biol.* (in the press).

Bubbling down

AIR-bubbles in water are usually rather deformed spheres. Daedalus reckons that toroidal bubbles should also be possible. He sees them as a form of two-phase vortex-ring — a toroidal air-bubble stabilized by the centrifugal force of the toroidally spinning liquid around it. Toroidal bubbles should persist and propagate like liquid vortices, except that the gas inside them would make them buoyant. Armed with these predictions, DREADCO's physicists are injecting air into variously spinning liquids, hoping to make toroidal bubbles and study their hydrodynamics.

Daedalus's interest is far from academic. He recalls that you can keep a hamster under water, in an air-filled silicone-rubber tent. The rubber is so permeable to oxygen and carbon dioxide that the hamster can breathe dissolved oxygen evolved from the water, and dispose of its expired carbon dioxide by the same route. A free water surface would, of course, be an even better gas-exchange membrane. Accordingly, says Daedalus, people could live and work under water in a big bubble, if it could be stabilized somehow. A spin-stabilized toroidal vortex-bubble should be ideal.

So Daedalus is planning a toroidal bubble-submarine, or 'Bubsub'. It is basically a ring-shaped raft, launched by an ingenious hydrodynamic wind machine which pushes it down into the water while blowing a toroidal bubble around it and stirring up a suitable spin. As their craft is forced powerfully downwards, the crew will see themselves magically enclosed in a sort of air-filled doughnut surrounded by spinning water. Once well submerged, the Bubsub will be stabilized by its own motor, whose drive shaft extends through the inner wall of water and through a right-angled gear drive to an upthrusting central propeller. The propeller's powerful upflow maintains the toroidal vortex, and also pushes the Bubsub down against its inherent buoyancy. Both crew and motor breathe normally, through the huge extended surface of the bubble, and will have a wonderful view of the marine world around them.

The Bubsub will be ideal for underwater exploration of all kinds. Biologists with butterfly nets will reach down through its walls to trawl plants from the sea bottom, or out to capture fish swimming past it (indeed, some may swim straight through its walls and fall astonished on the deck). Tourists will navigate scenic marine grottos and oceanographers search for sunken wrecks. The Bubsub must never be allowed actually to touch the bottom or any other solid surface, or the vortex would be fatally interrupted. But at the end of a voyage it can be allowed to float to the surface. The bubble will burst in a cloud of spray, leaving the crew safely afloat on the raft to await their mother-ship. David Jones

Cleaning up gene databases

SIR—We have discovered errors in the EMBL nucleotide sequence databank in the course of training of artificial neural networks to recognize pre-messenger RNA splicing signals¹ in human genes. In training on thirty-three human genes, the seven genes listed in the table appeared to disturb the learning process. Further

randomly lower the degree of regularity, which is also a prerequisite for the function of a specific biological recognition mechanism.

The problem of errors in the sequences incorporated in databanks could be amplified in the future, as a consequence of the increasing rate at which sequences are

Errors in the EMBL nucleotide sequence database

Databank sequence (Accession no.)	Published sequence	Type of error
HSHLIA (X00492)	ref. 5	Splice donor cited at base pair (bp) 328 in feature table (FT). Correct position: 329 (there is a typographical ambiguity in the article)
HSBGL3(V00499)	ref. 6	Splice donor cited at bp 2971 in FT. Correct position 2981
HSERPG (X02158)	ref. 7	Presumed -1 bp error in splicing frame of intron I in article* Nucleotides CG missing after bp 578 All following FT entries -2 bp wrong until last acceptor site which is +3 bp wrong
HSDGL1(V00505)	ref. 8	Presumed +1 bp error in splicing frame of intron I in article*
HSEGL1(V00510)	ref. 9	Presumed +1 bp error in splicing frame of intron I in article*
HSHSP27(X03900)	ref. 10	Presumed -1 bp error in splicing frame of intron I in article*
HSGROW2(V00520)	ref. 11	Splice donor cited at position 1081 in FT. Correct position: 1091

Discrepancies between published and databank entered sequences, or presumed errors in the published sequences.

* The occurrence of repeated bases around splice donors and splice acceptors means that the splicing frame of an intron cannot always be unambiguously assigned on the basis of the messenger RNA sequence and amino-acid sequence. A neural network can, however, point out the frame by recognition of the splice signals present in the sequence. Splicing always occurs at a position that is fixed with respect to the sequence signal¹².

investigation revealed discrepancies between the databank sequences and the original published papers for three of the genes. In the other four examples, we found wrongly assigned splicing frames of introns.

The subset of thirty-three human genes from the EMBL databank carried information about the location of splicing signals. On the basis of the feature table information, we assigned each nucleotide to one of two categories, splicing donor sites or otherwise. We used these data to train a neural network of the perceptron type² to classify nucleotides, on the basis of their context, as belonging to one of the two categories. The network was similar to that used for predicting secondary structures of proteins^{3,4}.

Neural networks are now being applied to many classification tasks; they have in common the ability, when trained, of dealing with non-linearities in the association between objects and categories. But in the presence of strong non-linearities, as when many similar objects must be put into widely different categories while many dissimilar objects are destined for the same category, the classification can usually be learned by the network only with difficulty. Errors introduced

determined and by the initiative to sequence the human genome. We therefore propose that computerized proof reading should be incorporated in the databanks, so as to take advantage of the ability of neural networks to reveal non-linearities in a dataset, as we have demonstrated.

SØREN BRUNAK

Department of Structural Properties of Materials,
The Technical University of Denmark,
DK-2800 Lyngby,
Denmark

JACOB ENGELBRECHT

Department of Dairy Science,
Royal Veterinary and Agricultural University,
Bülowsvej 13,
DK-1870, Frederiksberg C,
Denmark

STEEN KNUDSEN

University Institute of Microbiology,
Øster Farimagsgade 2A,
DK-1353 Copenhagen C,
Denmark

1. Padgett, R.A., Grabowski, P.J., Konarska, M.M., Sheller, S. & Sharp, P.A. *A. Rev. Biochem.* **55**, 1119–1150 (1986).
2. Rumelhart, D.E., Hinton, G.E. & Williams, R.J. in *Parallel Distributed Processing* (eds McClelland, J.L., Rumelhart, D.E. and the PDP Research Group) 318–362 (MIT, Cambridge, Massachusetts, 1987).

3. Quian, N. & Sejnowski, T.J. *J. molec. Biol.* **202**, 865–884 (1988).
4. Bohr, H. *et al.* *FEBS Lett.* **241**, 223–228 (1988).
5. Strachan, T., Sodoier, R., Damotte, M. & Jordan, B.R. *EMBO J.* **3**, 887–894 (1984).
6. Lawn, R.M., Efstratiadis, A., O'Connell, C. & Maniatis, T. *Cell* **21**, 647–651 (1980).
7. Jacobs, K. *et al.* *Nature* **313**, 806–810 (1985).
8. Spritz, R.A., DeRiel, J.K., Forget, B.G. & Weissman, S.M. *Cell* **21**, 639–646 (1980).
9. Baralle, F.E., Shoulders, C.C. & Proudfoot, N.J. *Cell* **21**, 621–626 (1980).
10. Hickey, E. *et al.* *Nucleic Acids Res.* **14**, 4127–4145 (1986).
11. DeNoto, F.M., Moore, D.D. & Goodman, H.W. *Nucleic Acids Res.* **9**, 3719–3730 (1981).
12. Jacob, M. & Gallinaro, H. *Nucleic Acids Res.* **17**, 2159–2175 (1989).

Neutron irradiation of superconductors

SIR—We were interested to see additional technical work¹ concerning the effect of neutron irradiation on critical currents in high-temperature superconductors. Our first paper on single-crystal neutron irradiation in $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ demonstrated the improvements that were possible², and suggested additional work needed for the technique to reach its full potential^{3,4}. Our most recent work⁵ showed an enhancement of the critical current due to neutron irradiation which increases strongly with temperature and is still effective in magnetic fields up to 8 T.

The enhancement factors are comparable to those reported in ref. 1. However, these factors are not in themselves very meaningful: if the critical current density of a very high-quality ($J_c=0$) single crystal is compared to the irradiated state, the enhancement factor will be infinity!

If comparisons of this kind are to be made, it is necessary to measure the critical currents before and after neutron irradiation on the same single crystal, as we have been careful to do in our work. In addition, significant flux motion effects must be included in the discussion, especially when results obtained by different measuring techniques are compared. These effects were ignored in ref. 1.

GEORGE W. CRABTREE

Argonne National Laboratory,
Argonne,
Illinois 60439, USA

HAROLD W. WEBER

Atomic Institute of the Austrian Universities,
Schuttelstrasse 115,
A-1020 Vienna, Austria

1. van Dover, R.B. *et al.* *Nature* **342**, 55–57 (1989).
2. Umezawa, A. *et al.* *Phys. Rev.* **B36**, 7151–7154 (1987).
3. Weber, H.W., Crabtree, G.W., Umezawa, A., Lju, J.Z. & Nunez, L.H. in *High-T_c Superconductors* (ed. Weber, H.W.) 273–278 (Plenum, New York, 1988).
4. Weber, H.W. *et al.* in *MRS International Meeting on Advanced Materials* Vol. 6 (eds Tachikawa, K. & Kitazawa, K.) 365–369 (Materials Research Society, 1989).
5. Weber, H.W. in *Studies of High Temperature Superconductors* Vol. 3 (ed. Narlikar, A.V.) 197–227 (Nova, New York, 1989).
6. Sauerzopf, F.M., Wiesinger, H., Weber, H.W., Crabtree, G.W. & Liu, J.Z. *Poster presented at Conf. on Materials and Mechanisms of Superconductivity/High-Temperature Superconductors*, Stanford, 23–28 July (1989); *Physica C* (in the press).

Persistent Arctic ozone layer

SIR—The Arctic ozone layer has been studied since 1935 from the Norwegian station in Tromsø (70° N). Long series of observations are also available for the stations in Oslo (60° N) and Spitsbergen (79° N). The Oslo station, where the measurements are not hampered by the

Figure 1 shows that the 30-hPa (~23-km altitude) temperature was sufficiently low for a few days in the period during which we observed the mini-hole (shaded area), and at this latitude there is sunlight for a few hours every day. Based on balloon experiments (Kiruna, Sweden) in January

1989, Hofmann *et al.*¹ suggested a small ozone destruction due to a combination of PSC formation and sunlight. The Oslo data do not conflict with this.

The ozone values observed at our stations were above average from early February 1989 (Figs 1, 2). Thus, the spring value (average of all measurements in the three months February, March and April) for Oslo was 6.5% above average and in Tromsø 10.5% above average (the highest value in ~45 years).

The ozone measurements from the Norwegian stations during the first four months of 1989 exhibit periods with both record low and high values. For trend analyses, it is therefore necessary to use the average value for a well

defined time interval such as a month, a season or a year. In Fig. 2 the spring values for Tromsø (back to 1936) and Oslo are given. In Fig. 3 the summer values (average of the measurements in May, June, July and August) and the winter values (average of December, January, February and March) for Oslo (1978–1989) are presented together with Bojkov's² results for 12 ozone stations north of 59° N for

excellent agreement with the data² for 12 different stations for the period in which they overlap (1978–1986). This observation indicates that the Oslo station is representative for the Arctic; (3) the data for

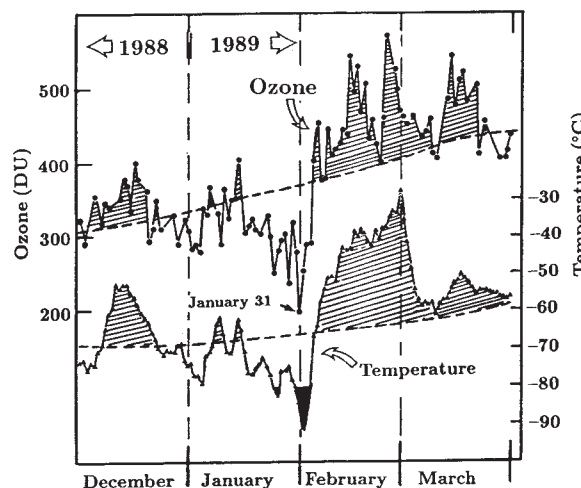


FIG. 1 Daily measurements of ozone (in DU) in Oslo and the 30-hPa (~23-km altitude) temperature (ref. 3) for the position 60° N, 10° E (slightly west of Oslo) during the winter 1988–1989. The dashed lines are the average values. Periods above average are hatched. The shaded area indicates that the temperature was below -80°C in the period with the large ozone depletion.

polar night, seems to be representative of the northern region.

Long-term measurements are needed for trend analyses, whereas the daily measurements give information about episodes with unusual ozone values. The data from long-term ozone measurements reveal periods of several years with a negative trend and other periods with a positive trend. The combined results up to 1989 give no evidence for a long-term negative trend of the Arctic ozone layer.

During the last days of January 1989 the ground-based Norwegian stations, as well as Nimbus 7, observed a 'mini-hole' in the ozone layer, centred over Scandinavia. On 31 January (Fig. 1), an ozone value of only 200 DU (dobson units) about 45% below normal, was observed by the Oslo station. The depletion lasted for a few days, followed by a rapid increase and record high values early in February. According to satellite data and 100-hPa (~16-km altitude) temperature contours, the 'mini-hole' was apparently filled with ozone-rich air from Canada and Alaska.

The ozone-depletion episode over Scandinavia was probably of meteorological origin, although we cannot exclude a small contribution due to heterogeneous chlorine chemistry involving chlorofluorocarbons (CFCs). The conditions for this type of ozone destruction include both low stratospheric temperatures (below $\sim -80^{\circ}\text{C}$) with the formation of polar stratospheric clouds (PSCs) as well as sunlight.

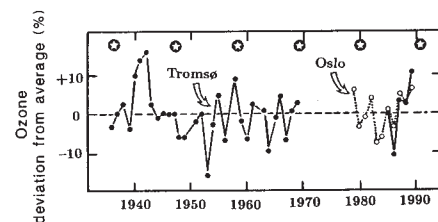


FIG. 2 Spring values (average of measurements in February, March and April corresponding to August, September and October in Antarctica) of ozone for Tromsø (filled circles) and Oslo (open circles). The data are given as per cent deviation from the average value, which for Oslo is 415 DU. For Tromsø we have a long period with no measurements. The stars indicate sunspot maxima. The Oslo data may indicate a correlation with the sunspot activity.

the period 1957–1986. The following conclusions can be drawn from the results presented in the two figures: (1) the ozone concentrations through the summer exhibit a smaller annual variation than both the winter values and the spring values; (2) the measurements for Oslo are in

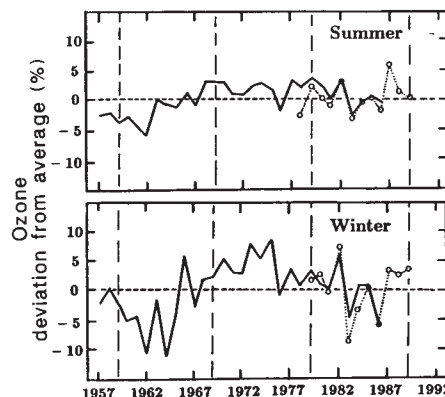


FIG. 3 Ozone data for Oslo for the period 1978 to 1989 and Bojkov's data² for 12 stations north of 59° N for the period 1957 to 1986. Summer data are the average of measurements in May, June, July and August and winter data the average of measurements in December, January, February and March. The results are given as per cent deviation from the average value which for Oslo is 368 DU (summer) and 373 DU (winter). The vertical lines indicate ten-year periods.

Oslo and Tromsø show that the ozone layer over Scandinavia has been above normal (or average) during the past three years. Because of the good correlations with the data from other stations this conclusion may be valid for the whole Arctic region; (4) the figures show the importance of defining the starting point and endpoint when describing trends. The data indicate a positive trend for ozone (in all seasons) in the period 1983–1989 (the past six years). On the other hand, no particular trend can be claimed for the past ten years. The 'winter data' in Fig. 3 indicate a small negative trend over the past 20 years and a similar, but positive, trend for the past 30 years; (5) These data indicate that anthropogenic gases such as CFCs have, up to the summer of 1989, had a negligible influence on the Arctic ozone layer. The general balance between formation and destruction of ozone has not changed, at least not to an extent that is apparent in the long-term observations. Whether heterogeneous chlorine chemistry contributes to transient ozone depletions such as the mini-hole observed over Scandinavia in the last days of January 1989 is not yet established, but the possibility should be considered.

SØREN H. H. LARSEN
THORMOD HENRIKSEN

*Institute of Physics,
University of Oslo,
Blindern, 0316 Oslo 3, Norway*

- Hofmann, D. J. *et al.* *Nature* **340**, 117–121 (1989).
- Bojkov, R. D. *Met. atmos. Phys.* **38**, 117–127 (1988).
- Naujokat, B., Labitzke, K., Lenschow, R., Petzoldt, K. & Wohlfart, R. C. *Beilage zur Berliner Wetterkarte* **SO 12** (1989).

Establishing a bench mark

Peter Harman

The Uses of Experiment. Studies in the Natural Sciences. Edited by David Gooding, Trevor Pinch and Simon Schaffer. Cambridge University Press: 1989. Pp. 48. Hbk £45, \$80; pbk £17.50, \$29.95.

The Development of the Laboratory. Essays on the Place of Experiment in Industrial Civilisation. Edited by Frank A. J. L. James. Macmillan Press, Houndsmills, Basingstoke: 1989. Pp. 260. £37.50.

THE publication of these two collections, both having their origins in recent conferences, reflects the current interest among historians of science in the role and status of experiment and the organization of scientific inquiry. Much emphasis is given, especially by the editors of *The Uses of Experiment* and by some of their contributors, to another focus of concern in the work of some (but by no means all) historians of science: the interpretation of scientific knowledge itself (and not merely the forms in which scientific research is organized) in trenchantly sociological terms.

Although the focus of the two volumes is different, three authors contribute to both collections; cross-reference to unpublished papers and books written by the various contributors to the volumes abound. The mutual self-congratulation by the editors and by some of the authors ultimately cloy, and although the editors of *The Uses of Experiment* are at pains to deny charges of "modish ignorance and neophilia", their own practice is not free from blame in this respect.

The editor of *The Development of the Laboratory* has collected thirteen papers, not all of them bearing directly on the stated theme of the book. He groups them into three sections: on chemical laboratories, on the extension of laboratories to physics, and on large physics laboratories. The second group contains papers most closely related in theme, for several contributors focus on the organization of laboratories, institutes and observatories in the nineteenth century. These studies, although rather specialized, do provide new information about laboratory design and manpower. The papers are written at varying levels of detail and sophistication, but provide some introduction to the styles of laboratory physics of the period (including the importance of metrology), and of the role of laboratories in physics teaching. The last part of the book collects papers

dealing with aspects of contemporary 'big science' (such as CERN and Fermilab), which should be of interest to physicists, especially those who work in these organizations.

The volume as a whole, however, illustrates the difficulty attendant on any

terizing the nature of the experimental process. The actual topics tackled by the contributors are wide and varied, and the authors write from different points of view (including, in the words of the editors, "historical epistemology, constructivist sociology, and literary criticism"), but the volume as a whole sustains considerable unity of purpose.

It does not always make for easy reading, but fortunately the editors' predilection for dense slabs of sociological prose is not shared by all the contributors. Including papers by philosophers and sociologists, as well as by historians of science (though these last are generally heavily sociological in their method), *The Uses of Experiment* will generate much interest and debate among students of

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Putting it to the test — the discovery of potassium and sodium by electrical decomposition, Humphry Davy (1807).

attempt to enclose the proceedings of a variegated conference within hard covers. The papers establish that there is indeed a subject to be investigated — denoted by the grandiloquent subtitle of the volume — but (hardly surprisingly) the book does not satisfactorily describe it.

The Uses of Experiment is a much more ambitious undertaking. As in the case of *The Development of the Laboratory*, the papers were selected from conference proceedings and were subsequently augmented by additional contributions; but considerable effort seems to have been expended in producing a volume of much greater unity and intellectual interest. The papers are grouped into five sections: the role of instruments in experiment, the rhetoric of experiment, the representation of experimental data, the audience for experiment, and the problem of charac-

science studies.

The book opens with an excellent account of eighteenth-century electrical instruments and apparatus by W. D. Hackmann. He is surely correct in arguing that the roles of scientific instruments and techniques are generally underestimated in accounts of the development of science. The study of contemporary apparatus, particularly of extant artefacts, should be central to any study of the role of experiment. In an essay on Newton's prism experiments, always a classic topic in any discussion of experimental procedure and argument, Simon Schaffer emphasizes the attempts to make Newton's famous 'crucial experiment' authoritative: experimenters' evaluations of the quality and arrangement of their prisms meshed with the definition of a 'good' experiment as one which produced the result that Newton claimed

to establish. Schaffer traces the changing fortunes of the prism experiments in an attempt to understand the nature of experimental authority.

The way a scientist seeks to articulate and represent experimental phenomena is explored in an interesting paper by Peter Galison and Alexei Assmus on C. T. R. Wilson's famous cloud chamber (the earliest particle detector), a classic piece of detection apparatus which has been used to give concrete meaning to the subatomic world of the twentieth-century physicist. Galison and Assmus emphasize that the origins of the device lie in Wilson's interest in the replication of cloud formations in the laboratory, arguing that Wilson's work began in a tradition of "mimetic experimentation", the attempt to reproduce natural physical phenomena in the laboratory.

Together with metrology and instrumentation, mimesis emerges as an important theme in the history of experimentation. David Gooding's essay on Faraday's magnetic curves, emphasizing the practicalities and particularities of experiment, "the seamless web of practical, intellectual and social interactions" that shaped Faraday's work, contrasts oddly with his paper in *The Development of the Laboratory*, where he reports a repetition of some of Faraday's classic experiments without giving any details of the procedure followed or the materials used.

J. A. Secord examines the controversial claim by Andrew Crosse to have created insects in his essay on the role of the audience for science in early Victorian England. The excitement generated by Crosse's claim tells much about the scientific community of the period and the new audience for science, and Secord's paper is one of several current investigations of the roles of audience and authority in the relations between popular culture and the culture of science in Victorian England.

The essays by Schaffer, on Newton, and by Secord thus complement one another in discussing experimental science in terms of authority and consensus. Such questions invite sociological analysis, and it is this style of argument which dominates the editors' introduction to *The Uses of Experiment*. It is indeed the uses of experiments which are the main subject of discussion, rather than the process of experimentation. This latter topic is, as the editors rightly state, neither "unproblematic or uninteresting"; but its historical study requires more reference to the notebooks and published papers of scientists, and less emphasis on problems of rhetoric, authority and social context. □

Peter Harman is in the Department of History, University of Lancaster, Bailrigg, Lancaster LA1 4YG, UK.

Man's best friend

Stephen Oliver

Yeast Genetic Engineering. Edited by P. J. Barr, A. J. Brake and P. Valenzuela. Butterworths: 1989. Pp.376. £55, \$65.
Molecular and Cell Biology of Yeasts. Edited by E. F. Walton and G. T. Yarranton. Blackie, Glasgow; Van Nostrand Reinhold, New York: 1989. Pp.354. \$112.95.

YEAST has been a friend to man for at least 8,000 years, providing him with his daily bread and beer or wine. It is also a favoured tool for molecular and cell biologists; if there could be a eukaryotic equivalent of *Escherichia coli*, in terms of experimental usefulness, then yeast is it. The wealth of scientific knowledge about yeast, when added to the millennia of practical experience with the organism, makes it the workhorse of modern biotechnology. The pleasure of working with yeast, as Barr and his colleagues point out, is that fundamental and applied research enjoy a synergistic relationship. This has been true since Pasteur solved the problems of the French brewers and, almost in passing, disproved the theory of the spontaneous generation of life.

The genome of budding yeast, *Saccharomyces cerevisiae*, can be manipulated with exquisite accuracy because of the organism's fastidious sexual habits — it has an absolute requirement for homology in recombination. Thus, the relationship between gene structure and function can be determined with great ease. *Yeast Genetic Engineering* provides a wealth of examples of how this can be exploited for both scientific and technological purposes.

Multi-author books are often the products of major symposia, but here, on the other hand, is a book which would have made a great scientific meeting. One can well imagine lively discussions between its contributors on such topics as protein secretion and gene regulation. As a book, it is rather disjointed and repetitive, so that even such a stalwart champion of yeast as myself tires of being told, in almost every chapter, of the organism's advantages as an experimental tool.

As industrial researchers, the editors are greatly concerned with the ability of yeast to secrete heterologous proteins, and at least six chapters address this problem to some extent. Degradation, an alternative fate for proteins after synthesis, is dealt with in a fascinating chapter by Varshavsky and his colleagues on the ubiquitin pathway. The role of messenger RNA structure in determining translational efficiency is handled by Clements *et al.* in a manner which does much to clarify a difficult area. Another contribution of admirable clarity is that from

Kingsman, Kingsman and Adams on yeast transposons. Their chapter, together with that of the editors, provides much hope for yeast becoming a vehicle for the production of safe and effective vaccines. The claims of non-*Saccharomyces* yeasts are not ignored, with Yamamoto describing the fission yeast, *Schizosaccharomyces pombe*, and Toh-E contributing an excellent chapter on the lactose-fermenting yeast, *Kluyveromyces*. Shuster provides a thoughtful, and much needed, industrial perspective.

In its first five chapters at least, *Molecular and Cell Biology of Yeasts* is much more logically organized and comprehensive in scope than *Yeast Genetic Engineering*. Mellor deals with the initiation of transcription and provides a sure guide through the maze of consensus sequences and activation factors involved in its regulation. The editors provide an object lesson in the construction of expression cassettes in a chapter on mating-type control, which contains much new information. In his contribution on messenger RNA translation and stability, Brown steers a course through such vexed areas as codon bias, translational control and messenger RNA turnover with sense and insight. A chapter by King and the two editors draws all this together in considering the production of heterologous proteins, while a comprehensive discourse on proteinases by Wolf and his colleagues provides an interesting new viewpoint on the secretion problem.

So far so good, but from here *Molecular and Cell Biology of Yeasts*, like *Yeast Genetic Engineering*, becomes something of a farrago. Hirado and Yanagida provide a chapter on the fission-yeast cell cycle, but it is odd that a volume devoted to molecular and cell biology should omit such topics as genome organization, chromosome structure and DNA replication.

Industrial interest is supplied by Hinchcliffe and Vakeria's account of the genetic manipulation of brewing yeasts. These authors discuss the prospects of engineering the production of human serum albumen to upgrade the value of spent brewery yeast. Human serum albumen is, perhaps, the only foreign protein that one might feel relaxed about finding in one's pint of beer. Nevertheless, it is evident that yeast's service to man is being rapidly extended from bread and beer to biomedical products, and that both these books give a sense of the excitement and confidence which pervades the field. □

Stephen Oliver is Director of the Manchester Biotechnology Centre, UMIST, PO Box 88, Manchester M60 1QD, UK.

■ *Yeast Biotechnology and Biocatalysis*, edited by H. Verachtert and R. De Mot, has just been published by Marcel Dekker. This volume details practical innovations in yeast production and growth and also examines promising developments in techniques. Price \$150.

Grains of truth

Peter D. Moore

Textbook of Pollen Analysis, 4th edn. Edited by Knut Fægri, Peter Emil Kaland and Knut Krzywinski. Wiley: 1989. Pp.328. £51, \$109.

HOUSED in the University of Cambridge Sub-Department of Quaternary Research are two boats used for the sampling of lake sediments; between them they support the platform from which the delicate yet strenuous operation of manual coring takes place. One boat is called *Fægri* and the other *Iversen*. One can hardly imagine a higher accolade, or a more fitting symbol of the way in which these two names have come to represent the support system of modern Quaternary palynology.

Fægri and Iversen first published their textbook in 1950, which in the following 25 years went through two further editions and some gradual evolution. The book rapidly assumed a unique position as a statement of the theoretical basis of pollen analysis and as a laboratory manual to assist in the practice of the technique. If the first three editions illustrate gradualism of evolution, this fourth edition undoubtedly displays saltation, taking on a new and hardly recognizable form. The book, like its subject, has evidently entered on a new lease of life.

Some aspects of the new edition remain the same, particularly its emphasis and coverage, which is essentially the Quaternary palynology of north-west Europe. Although many of its theoretical and methodological sections will be of value to those interested in other parts of the world and in other geological strata, the key to pollen identification remains firmly constrained within these limits. Apart from this, the entire approach has been changed and a welcome new vigour injected.

Worthy of particular mention are the chapters on the transport of modern pollen and the taphonomy of fossil pollen. Combined, these two chapters provide an excellent basis for the interpretation of pollen assemblages. Also new is a lengthy chapter on archaeopalynology — the use of pollen grains in archaeological contexts as a means of environmental reconstruction. Palynology has become an increasingly valuable tool for archaeologists, and this account should provide a useful guide to environmental archaeologists concerning the potential and the limitations of the technique.

Look at any of the old editions of Fægri and Iversen, sitting upon the shelves (or, more likely, the laboratory benches of practising palynologists), and you are sure to find that the most thumbed section will be the key to pollen identification. This is the part of the book that is guaranteed to be in daily use. Here, the most obvious change in the new edition is in the inclusion of small sketches in the body of the key illustrating important pollen features. These are extremely useful as guides to the user and are worth many words. Nevertheless, I still prefer photographs as it is so easy to idealize a diagram. I found it rather confusing that sometimes a small circle might represent an entire grain and sometimes a detail of pollen wall structure, for example. I certainly agree with the emphasis the authors place on the need to confirm identifications by reference to type material: there is no substitute for this discipline.

Sadly, Iversen has been dead for many years, but Fægri has made a splendid job of renovating this classic textbook, ably assisted by two colleagues at the University of Bergen who have bravely and very competently stepped into Iversen's shoes. □

Peter D. Moore is in the Division of Biosphere Sciences, King's College, Campden Hill Road, London W8 7AH, UK.

National debts

Nansen Behar

Environmental Policies in East and West. Edited by Gyorgy Enyedi, August J. Gijswijt and Barbara Rhode. Taylor Graham, 500 Chesham House, 150 Regent Street, London W1R 5FA/Suite 187, 12021 Wilshire Boulevard, Los Angeles: 1988. Pp.401. £35, \$65.

As WE approach the end of the century, the global ecological problem, to which pollution is one of the main contributors, continues to intensify. Pollution does not recognize national frontiers. To tackle it, international cooperation is essential.

Environmental Policies in East and West addresses the regional problems of Europe's environment and the policies of the individual countries concerned. The contributors have been guided by the recognition that pan-European environmental policy has to be determined both by natural variations from area to area (such as those in drainage basins and patterns of climate) and by differences in the intensity and siting of industry and agriculture. Beyond that, specific policy in each country in part depends upon the prevailing economic system — that is, on whether it has a market economy or a centrally planned economy. The overall premise on which the book is based is that individual nations have to become well acquainted not only with the environmental issues that concern their neighbours but with how policy is determined to tackle them. Only with that understanding can there be successful economic cooperation in environmental protection.

Brundtland's words from *Our Common Future* (p.xiii) are the apposite motto for the book: "The fact that we all became wiser, learnt to look across cultural and historical barriers, was essential. There were moments of deep concern and potential crisis, moments of gratitude and achievement, moments of success in building a common analysis and perspective." So, have we Europeans become truly wiser in our attitudes towards nature? Here, experts from 14 countries attempt to provide the answer through discussion of case-studies. Brought together under the auspices of the European Coordination Centre for Research and Documentation in Vienna, the authors are specialists in geography, sociology, economics and engineering among others — all disciplines with relevance to environmental matters. The choice of nations represented is well balanced, both geographically and socio-politically. From the West are Denmark, the Netherlands, Britain, France, West Germany, Italy and Austria; from the East, Bulgaria, Czechoslovakia, East Germany, Hungary, Poland, the



Probably one of the earliest drawings depicting a reforestation project, part A shows places where trees have been felled and only the stumps remain to be removed; B depicts a root being sawn into pieces; C–F describe the process of preparing the root for replanting on the hills. This drawing is reproduced from *A Forest Journey* by J. Perlin, published by W. W. Norton, £14.95, \$23.50.

Soviet Union and Yugoslavia. The case-studies have a more or less similar structure; each starts with a description of natural conditions in the country under consideration, moves on to economic and environmental factors, and the institutional framework, and ends with a discussion of decision-making processes. Comecon and EEC environmental policies are also taken into account.

Throughout the book there is a sense of concern that neither the market system nor the centrally planned system, as they stand, hold the answer as to how a balanced co-evolution between mankind and nature can be achieved alongside technological progress. A market economy induces artificial needs and does not always have the means of counteracting the interests of the big producers of goods. Planned economies, as yet, have not introduced effective ecological criteria for the direction of new investments; existing incentives are inadequate to make

environmental decision-making preventive rather than following events. The conclusion, implied by the contributors, is that Europeans need a new attitude towards the natural world and the complex balance between man, production and nature.

The book has some shortcomings. The quality of the case-studies is variable, some being more thorough and analytical than others. More cross-national comparisons would have been helpful — the effectiveness of any one country's environmental policy has to be judged not only in its own terms but against that of others. Nonetheless, the book is a pioneer work of comparative analysis in environmental decision-making. As such, environmentalists and administrators, as well as the large numbers of the public now taking a keen interest in ecological matters, will find it illuminating reading. □

Nansen Behar is in the Institute for Contemporary Social Studies, Bulgarian Academy of Sciences, 1635 Sofia, Bulgaria.

Probability test

David Miller

Scientific Reasoning: The Bayesian Approach. By Colin Howson & Peter Urbach. Open Court, La Salle, Illinois 61301, USA: 1989. Hbk \$34.95 £33.95; pbk \$16.95, £14.95. Distributed in Great Britain by Eurospan, 3 Henrietta Street, London WC2E 8LU.

IN 1970, the Italian probability theorist Bruno de Finetti estimated that it would take 50 years, or a bit more, for the ideas he embraced to oust classical statistics and become established in both Europe and the United States. Lindley paraphrased him by saying that "we shall all be Bayesian by 2020". Howson and Urbach here offer an approach to probability and statistics that differs in important ways from de Finetti's (they countenance objective physical probabilities, for example) but agrees on the main points.

Fundamental to the book is the view that, because the truth values of scientific hypotheses cannot be discussed with certainty, they should not be discussed at all; the most we can do is to evaluate the probabilities, given the evidence we have, of the hypotheses we have. Objective relations between these probabilities are imposed by the probability axioms, but the values are inescapably subjective; they are personal degrees of belief. The adjustment of probabilities on the receipt of new evidence is called inference, though it does not involve a passage from premises to conclusion. It is held that the characteristic features of scientific method can be properly understood only in these probabilistic terms.

In some cases — the explanation (p. 81) of the effect of refutations — the probabilistic gloss seems not exactly wrong, just silly. In others — the assertion (p. 110) that the only methodologically bad thing about *ad hoc* tinkering is that the resulting hypotheses have low probabilities — what is said is just wrong. But the main problem with Howson and Urbach's book is that it fails to make a scrap of sense of scientific activity itself. If only the probabilities of our theories matter, and not their truth or falsity, and if there is no sense in which an earlier state of knowledge (a prior probability distribution) is less correct than a later one (a posterior distribution), it is a mystery why experiments are ever conducted, or hypotheses submitted to test. This well-known objection (originally made by A. J. Ayer) is not faced by Howson and Urbach; and the well-known response to it (by I. J. Good), that a search for new evidence can sometimes be justified as an attempt to maximize expected utility, is inapplicable, as in the Bayesian picture of theoretical science no action follows on the receipt of new evidence and, indeed, there are no utilities. Of course, one could say (as the authors fleetingly suggest) that the point of empirical research is to raise (and so, necessarily, in some cases to lower) probabilities. That would be to concede that Bayesianism does not explain scientific method.

As well as promulgating their variant of Bayesianism, Howson and Urbach offer forthright criticisms of other theories of science — particularly those of Popper, Carnap, Fisher and other non-Bayesian statisticians. Some of these criticisms — for instance, that many procedures of classical statistics are forced to rely on judgments that are not obviously objective — are familiar. Others are shockingly in-

adequate: readers must not expect to find in section 1.1, entitled "Popper's Attempt to Solve the Problem of Induction", a fair description, let alone a fair criticism, of Popper's solution of the problem of induction. The authors would have done well too to have been more candid about Bayesianism's abdication of interest in matters of truth and falsity. For example, they criticize the classical theory of significance tests for sometimes issuing contradictory instructions on whether or not to reject the null hypothesis (p. 132); what they do not take the trouble to say explicitly is that the Bayesian solution involves giving no advice of any kind.

Perhaps the most dislikeable aspect of this book is the way in which the authors fail to direct against their own doctrines lines of criticism that they unsympathetically (and often inexpertly) direct against others. For example, Popper's demand that hypotheses admitted to science be empirically falsifiable, for example, is dismissed on the childish grounds that "an observation can only falsify a theory (that is, conclusively demonstrate its falsity) if it is itself conclusively certain", which statements based on evidence never are (p. 93). But when the question arises of how Bayesians cope with inconclusive evidence (p. 287), it transpires that all the treating of a statement as evidential means "is that the agent takes . . . [it] to be true in the light of his current experience", a decision in principle revisable. This sane advice does not differ much from what Popper proposed in 1934. I find it clear enough. But it must be incomprehensible to those who confess that "it is difficult to understand what opponents of the Bayesian approach could have in mind when they talk of theories being 'accepted' or 'retained', or 'put forward' . . ." (p. 106) and even (in the course of a criticism of Neyman-Pearson theory) devote a subsection to showing that they do not appreciate the distinction between accepting a hypothesis as true and accepting it as definitely true.

The book ends with a chapter that responds to a dozen or so of the purportedly most important objections to the Bayesian position (but not those raised above). I can only report that, although a few good points are made, I found these responses tendentious, evasive, needlessly rhetorical, and in not one case satisfactory.

De Finetti's forecast of the eventual dominance of Bayesian teaching was, remarkably, unqualified by any probabilistic hedge. Let us hope that — probable or not — the prediction turns out to be wildly incorrect. The appearance of this book warns us not to assume complacently that a philosophy of science so dreary and unenlightening cannot be adopted and enthusiastically advocated. □

David Miller is in the Department of Philosophy, University of Warwick, Coventry CV4 7AL, UK.

Impact delivery and erosion of planetary oceans in the early inner Solar System

Christopher F. Chyba

Laboratory for Planetary Studies, Cornell University, Ithaca, New York 14853, USA

The terrestrial planets may have acquired oceans of water (and other surface volatiles) as a late-accreting veneer from impacts of comets and carbonaceous asteroids during the period of heavy bombardment 4.5 to 3.5 Gyr ago. On any given body, the efficiency of this mechanism depended on a competition between impact delivery of new volatiles and impact erosion of those already present. For the larger worlds of the inner Solar System, this competition strongly favoured the net accumulation of planetary oceans.

BOTH carbonaceous asteroids¹ and comets² may have been important sources for Earth's surface volatile inventory, and in particular for its oceans. A long-standing objection to this hypothesis has been that large impacts might erode as much or more of the terrestrial volatile inventory as they deliver³. Analytical fits to the lunar cratering record, appropriately scaled to Earth's larger gravitational cross-section, allow estimates of the terrestrial bombardment history. By convolving these estimates with quantitative models for impact erosion of planetary atmospheres⁴ and condensed oceans, I demonstrate here that Earth probably accreted a net (after impact erosion) amount of water corresponding to ≥ 0.2 – 0.7 ocean masses during the period of heavy bombardment 4.5 to 3.5 Gyr BP (before present) (where the terrestrial ocean mass M_{oceans} is 1.4×10^{21} kg H₂O). Venus should have collected approximately as much water as Earth, whereas Mars, more subject to impact erosion, should have accreted an ocean equivalent to a layer ~ 10 – 100 metres deep distributed over the planet.

Ironically, the greatest uncertainty in these results arises from the part of the calculation that is based most strongly on empirical data. Differing lunar crater counts, as well as disparate acceptable fits to even the same lunar cratering data sets, lead to cumulative bombardment history models that differ by $\sim 10^3$ when extrapolated back to 4.5 Gyr BP. (Lunar geochemical data do allow a firm lower limit to be placed on total incident impactor mass, however, which yields the 0.2– $0.7 M_{\text{oceans}}$ cited above.) I also conclude that calculations of the impact frustration of the origins of life^{5,6} that rely on such bombardment models may suffer from correspondingly great uncertainties.

Uncertainties in the lunar cratering record

Attempts to estimate the impact environment of early Earth typically begin with an analytical fit to the lunar cratering record. One such fit is that found by Melosh and Vickery⁴ in their recent work on impact erosion of atmospheres. They give

$$N(>4 \text{ km}, t) = 2.68 \times 10^{-5} [t + 4.57 \times 10^{-7} (e^{t/\tau_a} - 1)] \text{ km}^{-2} \quad (1a)$$

as an analytical fit to empirical data (from the Basaltic Volcanism Study Project (BVSP, ref. 7, Table 8.4.2)) relating surface density of lunar craters with diameters $D > 4$ km, $N(>4 \text{ km})$, to the radiogenic age t of the cratered surface. Melosh and Vickery chose the time constant $\tau_a = 220$ Myr. The BVSP data, and the fit given by equation (1a), are shown in Fig. 1.

It is not clear that $\tau_a = 220$ Myr is the correct time constant

for the decay of the early impactor flux. Other authors have made different choices. For example, Maher and Stevenson⁵ modelled the impactor flux with $\tau = 70$ Myr, whereas Neukum⁸ favours $\tau = 144$ Myr (corresponding to a 100-Myr half-life). Fundamentally, imprecision is inevitable because, as has been emphasized^{9–11}, the impactor flux cannot actually have decayed at a constant rate. Rather, the flux must initially have been dominated by objects in nearly circular Earth-like orbits (for which $\tau \sim 20$ Myr); as this population was swiftly swept up, different, more slowly-decaying populations (with $\tau \sim 100$ – 300 Myr, such as comets or Mars-exchanged asteroids) would have become predominant. Thus fitting a decay 'constant' to the lunar data is a crude approximation at best. Nevertheless, the actual data (~ 15 points) may not justify a more sophisticated treatment (see, however, ref. 11, for such an attempt).

To demonstrate the range of decay rates permitted by the BVSP data, I have fit these data, using two-dimensional χ^2 minimization, with an equation of the form (1a) using $\tau = 144$ Myr. I have chosen this decay constant because a 100-Myr half-life has been demonstrated for the decrease of the primordial comet flux through the inner Solar System by independent numerical simulations of the formation of Uranus and Neptune^{12,13}. My fit, shown in Fig. 1 as the solid curve, is given by

$$N(>4 \text{ km}, t) = 3.5 \times 10^{-5} [t + 2.3 \times 10^{-11} (e^{t/\tau_b} - 1)] \text{ km}^{-2} \quad (1b)$$

where $\tau_b = 144$ Myr. Both equations (1a) and (1b) provide moderately good fits to the BVSP data, with χ^2 values of 8.06 and 10.6, respectively (for $15 - 2 = 13$ degrees of freedom). Equation (1a) is formally a better fit by $\sim 25\%$, but this comparison is of little significance. First, as just discussed, fitting any single decay 'constant' to the cratering flux is a procrustean exercise. Second, the largest discrepancy between the two fits occurs for the point at ~ 4.4 Gyr in Fig. 1. There is, however, no true radiogenic data for this point; rather its age is simply an estimate⁷. Furthermore, if the lunar uplands are saturation cratered (see below), the crater density plotted for this point (which is an uplands value) could be a considerable underestimate.

Recent work on the impact frustration of the origins of life^{5,6} uses a cumulative lunar crater density given by

$$N(>4 \text{ km}, t) = 1.4 \times 10^{-5} [t + 5.6 \times 10^{-23} (e^{t/\tau_c} - 1)] \text{ km}^{-2} \quad (1c)$$

with $\tau_c = 70$ Myr. Equation (1c) provides a reasonable fit to the lunar cratering data (for $D > 20$ km) given by Wilhelms (ref. 14, Fig. 6.49). Maher and Stevenson⁵ adopt a number-diameter scaling law $N \propto D^{-1.8}$ for lunar cratering, which they consider to hold at least to diameters as small as $D = 1$ km. In equation (1c), I have used this scaling law to convert Wilhelm's data (and their fit) to $D = 4$ km so that it may be compared directly (Fig. 1, dotted line) with the BVSP results. (Melosh and Vickery⁴ implicitly adopt a -1.8 scaling as well, in their conversion from cumulative crater density to heavy-bombardment flux.) This conversion is of course perfectly consistent with Maher and Stevenson's model, but not quite with Wilhelm's data, which are based on combining crater counts at $D > 20$ km and $D > 1$ km using an empirical law (derived from Imbrium basin counts) that is not consistent with -1.8 scaling. The $D > 1$ km counts are drawn from Neukum^{8,14}, whose small-crater counting methodology is incompatible with that adopted by BVSP⁷. It is not surprising that the results given by equation (1c) differ so

markedly from those of equation (1a,b). Each model gives similar results for ages ~ 3.8 Gyr because the data at these dates are in rough agreement. (It is the requirement that the models approximately agree at ~ 3.8 Gyr, despite differing time constants, that leads to the vastly different coefficients, 10^{-7} – 10^{-23} , in equation (1a)–(1c).) At dates ≥ 3.8 Gyr, however, the models diverge: extrapolation back to 4.5 Gyr gives cumulative crater densities that differ by $\sim 10^3$.

Another question that arises in comparing the BVSP and Wilhelms data sets is the role of secondary craters. BVSP generally include only craters with $D > 2.8$ km, to minimize contamination by secondaries. Basin secondaries as large as $D \sim 20$ – 30 km may, however, substantially contaminate crater counts in the lunar highlands^{7,15}. Nevertheless, the slopes of the crater diameter distributions used by BVSP for the oldest lunar surfaces show no steepening at small D , indicating either that few secondaries contribute to these counts or that these surfaces are saturation cratered.

Early terrestrial impact environment

Given equation (1a, b, c) the total mass incident on Earth during heavy bombardment can be calculated. N in equation (1) scales as $D^{-1.8}$; this crater diameter scaling can be converted to impactor mass scaling using a mass–diameter equation. Following Melosh and Vickery⁴, I use Melosh's¹⁶ version of the Schmidt–Housen¹⁷ relation, assuming an average impact angle of 45° and a crater collapse factor c_f of 30% (ref. 18)

$$m_p = (0.11) \rho_t^{19/15} \rho_p^{-1/4} g^{21/25} v_p^{-5/3} \times (1.3/c_f)^{19/5} (\sin 45^\circ / \sin \theta)^{5/3} D^{19/5} \quad (2)$$

Here m_p is the mass of the projectile in kg, $\rho_t = 2,900 \text{ kg m}^{-3}$ is the lunar crustal density¹⁹, $\rho_p = 2,500 \text{ kg m}^{-3}$ is the density of a typical impacting asteroid²⁰, $g = 1.61 \text{ m s}^{-2}$ is the gravitational acceleration at the lunar surface, $v_p = 1.3 \times 10^4 \text{ m s}^{-1}$ is a typical impact velocity of asteroids with the Moon (discussed below), D is measured in metres and θ is the impact angle. Melosh and Vickery⁴ chose $c_f = 1.25$ which would increase m_p by $\sim 16\%$ over the values I obtain. Inserting equation (2) into the implicit $(4 \text{ km}/D)^{-1.8}$ scaling in equation (1) yields three equations for the cumulative number, $n(>m, t)$, of impactors with mass $> m$ that have been incident on a lunar surface of age t . The total mass, $M(t)$, incident in impactors with masses in the range m_1 to m_2 is then

$$M(t) = \int_{m_1}^{m_2} m [\partial n(>m, t) / \partial m] dm \quad (3)$$

which yields

$$M(t) = 0.76 [t + 4.57 \times 10^{-7} (e^{t/\tau_a} - 1)] m_2^{1-b} \text{ kg}^b \text{ km}^{-2} \quad (4a)$$

$$M(t) = 0.99 [t + 2.3 \times 10^{-11} (e^{t/\tau_b} - 1)] m_2^{1-b} \text{ kg}^b \text{ km}^{-2} \quad (4b)$$

$$M(t) = 0.40 [t + 5.6 \times 10^{-23} (e^{t/\tau_c} - 1)] m_2^{1-b} \text{ kg}^b \text{ km}^{-2} \quad (4c)$$

for equation (1a)–(1c), where $b = (1.8/3.8) \approx 0.47$, t is in Gyr and I assume $m_2 \gg m_1$.

What is the appropriate value in equation (4) for m_2 ? An N – D power law for lunar cratering is well obeyed in both the lunar frontside highlands and heavily cratered uplands (that is, the oldest lunar surfaces) for histogram bins in D up to $D = 512\sqrt{2} \text{ km} = 724 \text{ km}$ (ref. 7, Figs 23, 26, section 8.11). $D = 724 \text{ km}$ in equation (2) corresponds to $m_p = 1.5 \times 10^{18} \text{ kg}$, which I use for m_2 in equation (4). It is true that larger basins exist on the Moon (see, for example, ref. 14, Table 6.4), but the statistics of the formation of such basins as a function of time are problematic. Thus the calculation of $M(t)$ in equation (4) is a conservative underestimate. Because $M(t) \sim D^2$, the effect of a different choice for the largest basin included in equation (4) is easily deduced.

To calculate the total mass incident on Earth since $t = 4.5$ Gyr BP, I multiply $M(t)$ by the lunar surface area and scale to the

larger gravitational cross-section of Earth. (4.5 Gyr has been chosen because the possible Moon-forming impact of a Mars-sized object with Earth is typically dated¹⁶ at ~ 4.5 Gyr and volatiles previously accreted by the outer layers of Earth would probably be lost in this event. The oldest lunar rocks have ages consistent with crystallization 4.5–4.6 Gyr BP, ref. 21.) A body's gravitational cross-section σ is $\pi r^2 [1 + v_{\text{esc}}^2/v_\infty^2]$ where r is the body's physical radius, v_{esc} is its escape velocity and v_∞ is the relative velocity at infinity of the approaching impactor. As discussed below, I use 17 km s^{-1} as a typical asteroidal impact velocity with Earth, corresponding to $v_\infty \approx 13 \text{ km s}^{-1} \approx v_p$. For $v_\infty = 13 \text{ km s}^{-1}$ the gravitational cross-section of the Earth is ~ 23 times bigger than that of the Moon. But, ~ 4.5 Gyr BP, the Moon was much closer to Earth than now. In this case, gravitational focusing by Earth leads to both an increased lunar gravitational cross-section (v_{esc}^2 is replaced with the sum of the squares of the lunar escape velocity and the escape velocity of Earth at the Moon's orbital radius, which decreases Σ , the terrestrial/lunar gravitational cross-section ratio) and a larger average lunar impact velocity (v_p in equation (2)). Dynamical calculations of the evolution of the lunar inclination may exclude accretion of the Moon in Earth's equatorial plane within $\sim 10 R_\oplus$ (Earth radii)²². At $10 R_\oplus$, the effects of terrestrial gravitational focusing on both Σ and v_p are $< 10\%$, negligible compared to other uncertainties in the problem. If the Moon did in fact form near the terrestrial Roche limit, $\sim 3 R_\oplus$, its orbital evolution out to $\sim 10 R_\oplus$ was almost certainly so rapid ($< 10^{-5}$ of the time required to evolve from 10 to $60 R_\oplus$, in a model with constant terrestrial specific tidal dissipation Q —see, for example, ref. 23, equation (8)) that $t \approx 4.5$ Gyr may be used as the time when the Moon was at $10 R_\oplus$.

I find that the total mass incident on Earth since 4.5 Gyr BP is $M_{\text{tot}} = 9.9 \times 10^{20} \text{ kg}$, $3.2 \times 10^{21} \text{ kg}$ or $6.9 \times 10^{23} \text{ kg}$, respectively,

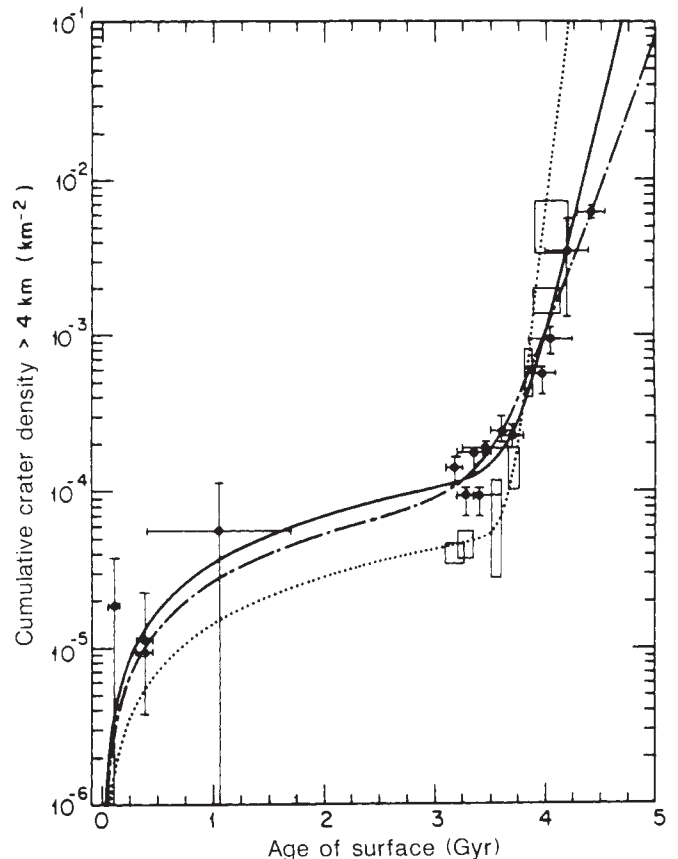


FIG. 1 Cumulative lunar crater density as a function of surface age, for the data in ref. 14 (boxes) and ref. 7 (crosses), fit by analytical formulae with 70 Myr (dotted line; ref. 5), 144 Myr (solid line) and 220 Myr (dot-dashed line; ref. 4), decay constants.

depending on the version of equation (4) used. This range is similar to the lower limit ($7.6 \times 10^{20} - 2.3 \times 10^{23}$ kg) found by summing the masses implied by all extant lunar craters and basins². By comparison, Earth's mass $M_{\oplus}$ is 5.98×10^{24} kg, and the terrestrial ocean mass M_{oceans} is 1.4×10^{21} kg H_2O .

These results may also be compared with a recent determination by Sleep *et al.*²⁴. They use iridium and nickel abundances of lunar highland rocks, together with estimates of the mixing depth of the meteoritic component of the lunar crust, to calculate a lower limit for the total amount of material striking the Moon since the solidification of the lunar crust ~ 4.4 Gyr BP. They conclude that the Moon has accreted an equivalent meteoritic thickness of 0.7 km, with an estimated uncertainty of a factor of two. Scaling to Earth with the parameters used here, and multiplying by a factor $\sim \exp[(4.5-4.4) \text{ Gyr}/100 \text{ Myr}]$, I obtain $M_{\text{tot}} = 4.8 \times 10^{21}$ kg as a lower limit for the total mass incident on Earth since 4.5 Gyr BP. Therefore the value $M_{\text{tot}} = 1.0 \times 10^{21}$ kg, derived from the 220-Myr fit to the lunar cratering record, is evidently an underestimate of the total impact flux. This conclusion is even stronger when one considers impact models in which the Moon would have lost much material in the ejecta²⁵ and vapour plumes⁴ resulting from high-velocity impacts. Although some of this material would have been trapped in Earth orbit and re-accreted by the Moon, much may have been lost from the Earth-Moon system altogether, because even as close as $10R_{\oplus}$, Earth's escape velocity is only 3.5 km s^{-1} , comparable to the Moon's (2.4 km s^{-1}).

Frustration of the origin of life

Certain conclusions of recent work on the impact frustration of the origin of life may be undermined by the uncertainties in the lunar (and hence, terrestrial) early impact environments described above. Maher and Stevenson⁵ have compared proposed timescales ($10^5 - 10^7$ yr) for the origin of life to timescales for globally sterilizing impacts, and argued that life could not have originated before 4.2 Gyr BP. Oberbeck and Fogleman⁶ corrected an error in the calculation and found that life could first have originated between 3.7 and 4.0 Gyr BP. Both these investigations, however, rely on the cratering time constant $\tau_c = 70$ Myr. Figure 1 indicates that their conclusions may depend strongly on this choice of τ .

$^{12}\text{C}/^{13}\text{C}$ isotope ratios in 3.8-Gyr Isua metasediments²⁶ suggest that life originated before 3.8 Gyr BP. It is clear from Fig. 1 that, for times ≤ 3.9 Gyr BP, all three fits to the lunar cratering record provide approximately equal fluxes and hence must give about the same results for impact frustration (V. Oberbeck, personal communication). For times earlier than ~ 3.9 Gyr, however, conclusions drawn from the different models dramatically diverge. For example, at 4 Gyr, equation (1c) gives a cumulative cratering flux N_c which is about five times greater than that given by equation (1a), N_a . Timescales for devastating impacts will be proportional not to cumulative fluxes, but to differential fluxes. The relevant ratio is therefore $(\partial N_c / \partial t) / (\partial N_a / \partial t) \approx 5(\tau_a / \tau_c) \approx 16$. That is, the frequency of impacts at 4 Gyr for a $\tau_c = 70$ Myr model is about sixteen times greater than that for a $\tau_a = 220$ Myr model. The frequency of impacts sufficient to frustrate the origin of life at 4 Gyr in the τ_c model is only achieved in the τ_a model at a time Δt before 4 Gyr, where Δt is given by $\exp(\Delta t / 220 \text{ Myr}) \approx 16$, or $\Delta t \approx 610$ Myr—which is before terrestrial accretion occurred. That is, frustration of life at 4 Gyr in the τ_c model implies there is no time in the τ_a model at which frustration occurs. These uncertainties emphasize the advantage of impact-frustration models that calculate the annihilation of assumed-extant ecosystems²⁴, or give as a result (rather than assume) a timescale for the origin of life²⁷.

Impact erosion of the hydrosphere

Melosh and Vickery⁴ find that large impacts may cause atmosphere erosion when two criteria are met. The first is that the impactor must strike the planet at a velocity high enough for a

vapour plume to form and expand at a speed $> v_{\text{esc}}$. Second, the mass of the plume must exceed the air mass above the plane tangent to the impact. In effect, these criteria are a momentum balance requirement. Here I consider an impactor that satisfies these two criteria to be entirely lost as an escaping vapour plume from the target planet. Not only will that impactor's volatiles not contribute to the target world, but atmospheric mass will be lost as well. As shown below for the case of the oceans, those volatiles present in a condensed state before an eroding impact will be largely immune to such erosion, resulting in the removal of some planetary volatiles (such as noble gases, nitrogen and carbon as carbon dioxide) and leaving others relatively unaffected.

Melosh and Vickery⁴ found that the threshold impact velocity v_{min} for most of a vapour plume to exceed v_{esc} was $v_{\text{min}} = \sqrt{4(v_{\text{esc}}^2 + 2H_{\text{vap}})}$, where H_{vap} is the vaporization energy, assumed to be 13 MJ kg^{-1} for silicates and 3 MJ kg^{-1} for ice. For the Moon, $v_{\text{min}} \approx 11 \text{ km s}^{-1}$ for silicate projectiles (asteroids) and $v_{\text{min}} \approx 7 \text{ km s}^{-1}$ for ice impactors (comets), so that virtually all vapour plumes resulting from cometary impacts (for comets of any mass, as there was no atmosphere to be overcome), and most resulting from asteroidal impacts, would be lost. This helps to explain why the Moon, although battered by carbonaceous asteroids and comets throughout its history, remains so volatile poor. For Earth, however, $v_{\text{min}} \approx 25 \text{ km s}^{-1}$ for asteroids, and $v_{\text{min}} \approx 23 \text{ km s}^{-1}$ for comets. These threshold values are analytical approximations to numerical work in preparation by Melosh and Vickery (personal communication); my results, however, do not depend strongly on the exact values chosen.

Short period (SP) comets have $v_{\text{rms}} = 28.9 \text{ km s}^{-1}$ (ref. 28), which has led to the suggestion³ that 'the idea that the Earth owes its oceans to cometary bombardment long ago may have to be abandoned'. Root-mean-square calculations of most-probable impact velocities assign greatest weight to those objects with exceptionally high values of v_{∞} , however, skewing the velocity distribution towards high velocities. In fact, many SP comets will hit Earth with velocities $< 23 \text{ km s}^{-1}$. I have used Weissman's²⁸ compilation of SP-comet velocities and collision probabilities to calculate the percentage of comet-Earth collisions that occur at a given velocity (Fig. 2). The statistics are poor, but Fig. 2 indicates that a small number of high-inclination and retrograde comets pull v_{rms} towards high values. In fact,

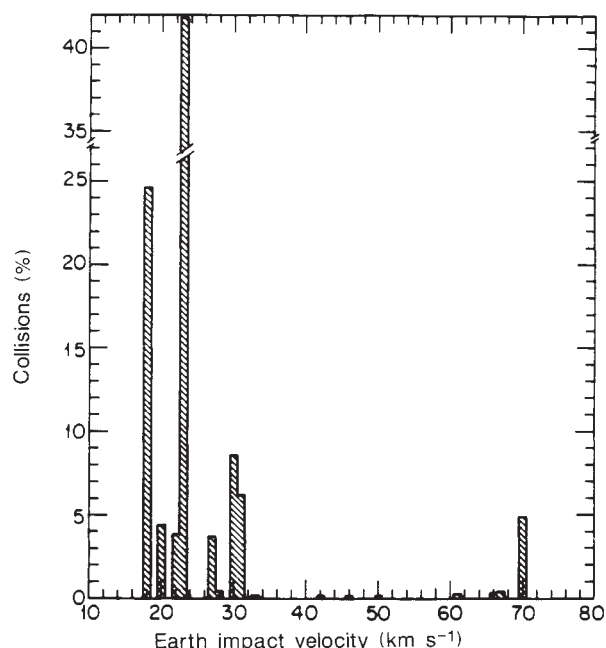


FIG. 2 The percentage of short-period-comet collisions with Earth as a function of impact velocity. Data from ref. 28.

~30% of comet-Earth collisions occur at or below 20 km s^{-1} and ~45% occur just at ~22–23 km s^{-1} (the critical velocity for escape of the resulting vapour plume). Assuming that half of the latter collisions cause atmosphere erosion, there remains a full ~50% of SP-comet collisions with Earth that do not create vapour plumes with sufficient velocity to reach escape speed. Because models of the early cometary bombardment of the inner Solar System^{12,13} indicate that the flux of comets scattered directly from the Uranus-Neptune region (that is, following SP-like orbits) dominated (by several orders of magnitude) the flux of those (long period) comets first scattered out to the Oort cloud, I consider that ~50% of the early comet flux incident on Earth had velocities low enough for their volatiles to have been retained. An analogous calculation using data for Earth-crossing asteroids (ref. 7, Table 8.5.1) shows that ~90% of asteroidal collisions with Earth occur at velocities $< 25 \text{ km s}^{-1}$. Yet these data also give a comet-like r.m.s. impact velocity for asteroids of ~25 km s^{-1} . Again, a small number of high-inclination objects skew the r.m.s. value. In fact, >50% of the terrestrial collisions of the Earth-crossing asteroids in this compilation will occur at velocities $< 17 \text{ km s}^{-1}$, the value chosen as 'typical' in the calculations above.

To determine the total mass of water accreted by Earth because of asteroidal and cometary impacts, I combine the calculation of M_{tot} above with an estimate of the division of M_{tot} into comet and asteroid fractions, with an estimate of the average water content of these bodies, and with the results of impact erosion in cometary and asteroidal collisions. The fraction α of the total ancient incident flux of comets is ultimately an unknown, free parameter. I have discussed elsewhere the variety of data pertaining to this question and argued that none is conclusive². It seems likely that $\alpha \approx 10$ –20%, but even this conclusion may be questioned. In the following discussion, I let α range from 10^{-3} to 1, and find that my results vary by less than a factor of four over this range.

Comets are ~50% water ice²⁹. It has been estimated²⁰, on the basis of photometric observations of Earth-crossing asteroids, that the asteroidal flux at Earth is an equal mix of S and C types. Only types C1 and C2 carbonaceous chondrite meteorites

match the spectra of C-type asteroids²⁰; I assume C-type asteroids to be an equal mix of these. C1 and C2 meteorites are ~20% and 13% water by mass, respectively³⁰. I assume that S-type asteroids contain no water.

Finally, could impactors erode a significant quantity of condensed water from Earth's surface? Consider an impactor with velocity and mass sufficient to cause atmosphere erosion. For a 1-bar terrestrial atmosphere, such an impactor has a minimum mass m_* of $3.5 \times 10^{15} \text{ kg}$, or a diameter $2r_p \approx 14 \text{ km}$ (ref. 4, equation (2)). The primordial terrestrial ocean would have accumulated its present surface-averaged depth $d_{\text{oc}} \approx 3 \text{ km}$; the following calculation maximizes impact erosion of this ocean by assigning it d_{oc} since 4.5 Gyr BP. An eroding impactor of diameter $\geq 2r_p$ deposits nearly all its impact energy at a depth $\sim 2r_p$; that is, it penetrates a 3-km ocean and continues many km into the sea bed¹⁶. The ocean water in the path of the impactor is rammed into the sea bed and severely shocked; this cylinder of water will enter the escaping vapour plume. For hypervelocity impact, the shock pressure in the ocean around this cylinder falls off as $\sim r^{-3}$ (ref. 16), so that little more water is sufficiently shocked to contribute to the escaping vapour plume (H. J. Melosh, personal communication)—of course, much more water is vaporized, but not lost from the planet.

Thus, the mass of ocean water eroded by an impactor of mass m , and diameter $2r_p$ is $M(m) = \pi r_p^2 d_{\text{oc}} \rho_{\text{oc}} = \pi (3/4 \pi \rho_p)^{2/3} m^{2/3} d_{\text{oc}} \rho_{\text{oc}}$, where ρ_{oc} and ρ_p are the densities of the ocean and impactor, respectively. The total mass $M_{\text{lost}}(t)$ of condensed ocean eroded by impactors of mass $> m_*$ is then

$$M_{\text{lost}}(t) = \int_{m_*}^{m_*} M(m) [\partial n(> m, t) / \partial m] dm \quad (5)$$

Performing this integration, and taking into account that only ~10% of asteroids and ~50% of comets have v_{min} large enough to cause erosion, I find that $\leq 5\%$ and $\leq 15\%$ of the water delivered by asteroids and comets, respectively, is subsequently eroded by incorporation of condensed water into vapour plumes.

The fractional volume of atmosphere removed by an eroding impact is $\sim 3 \times 10^{-4}$. With $\sim 10^{-5} M_{\text{oceans}}$ residing in the present atmosphere³¹, a calculation analogous to that above shows that loss by erosion of vapour-phase water is negligible. This result holds even in hypothesized 10-bar CO_2 early atmospheres, with ~0.5-bar H_2O in the vapour phase³². Finally, certain earlier models³³ of impact erosion indicated that erosion could not occur for worlds with $v_{\text{esc}} \geq 10 \text{ km s}^{-1}$; in this case Earth would have retained all incident volatiles.

Volatile inventories in the inner Solar System

Combining all of these considerations, the total amount of water accreted by Earth because of impacts over the past 4.5 Gyr is given by $M_{\text{H}_2\text{O}} = [(1 - \alpha)(0.95)(0.9)(0.5)(0.17) + \alpha(0.85)(0.5) \times (0.5)] M_{\text{tot}}$, where the first term is the asteroidal contribution (5% of condensed oceans eroded, 90% of impactors accreted, 50% type C, which are ~17% water by mass), and the second is cometary (15% of oceans eroded, 50% accretion, 50% water by mass). $M_{\text{H}_2\text{O}}$ scales erosion appropriately for the case $d_{\text{oc}} > 3 \text{ km}$ (as for equation (4c)). $M_{\text{H}_2\text{O}}$ is plotted as a function of α in Fig. 3, for M_{tot} calculated from both equation (4) and the empirical determination of Sleep *et al.*²⁴. Because this latter determination represents a lower limit, I conclude from Fig. 3 that Earth accreted a net (after impact erosion) amount of water ≥ 0.2 – $0.7 M_{\text{oceans}}$ during the period of heavy bombardment.

Even taking impact erosion into account, the cratering decay constant $\tau_c = 70 \text{ Myr}$ requires Earth to have accreted ~30–100 M_{oceans} of water since 4.5 Gyr BP. Barring creative, and *ad hoc*, mechanisms to bury this much water in Earth's mantle, this result indicates either that this choice of τ considerably overestimates the severity of the early terrestrial impactor flux, or that the assumptions used above regarding the water content of heavy-bombardment asteroids are incorrect. For example, if S types were in fact responsible for 90% of terrestrial collisions,

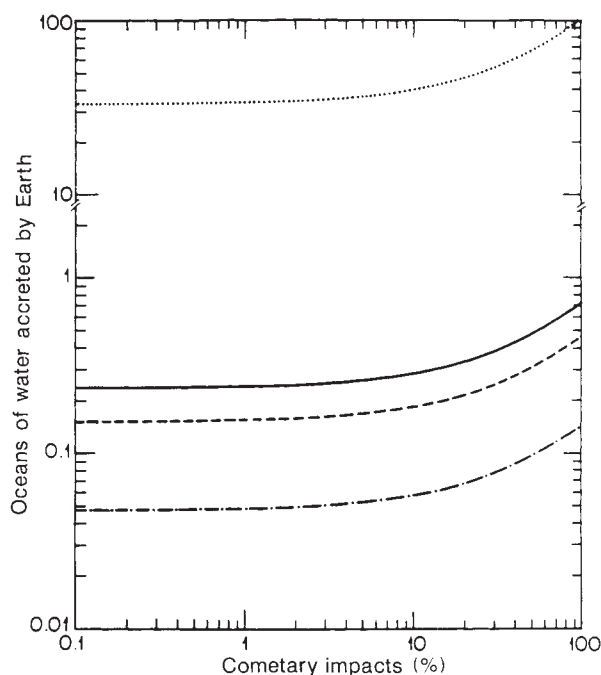


FIG. 3 Oceans of water accreted by Earth (in units of $M_{\text{oceans}} = 1.4 \times 10^{21} \text{ kg}$), as a function of the cometary percentage of impactor mass. Dot-dashed curve from equation (4a), dashed from equation (4b), dotted from equation (4c) and the solid curve is derived from ref. 24.

and if the cometary contribution were negligible, net water accretion would be reduced by a factor ~ 25 , so that $\tau_c = 70$ Myr would deliver only $\sim 1.2M_{\text{oceans}}$. It could also be that early in Solar System history, the impacting flux was dominated by water-poor objects in Earth-like orbits. The flux of such short-half-life objects would, however, have been superimposed over the flux of longer-lived objects represented by equation (1) (see for example ref. 12).

If α were as large as $\sim 20\%$, the water accreted by Earth would be dominated by comets. Is such a result consistent with the terrestrial inventory of other volatile elements, especially carbon and nitrogen? The elemental abundance ratios found for comet Halley²⁹ indicate that Earth, were its carbon primarily cometary in origin, should have about five times as much nitrogen as it actually has³¹. But both the preferential impact erosion of nitrogen and the likelihood that comets supplied only a fraction of the terrestrial volatile inventory defuse this dilemma.

Finally, these calculations have implications for the volatile

inventories of the other terrestrial planets. For reasons noted above, the Moon probably did not retain more than an insignificant fraction of the volatiles incident upon it (see ref. 2 for a further discussion). Delivery of volatiles to Venus and Mars depend on their impact rates relative to the Moon, but the analysis presented above indicates that Venus should have accreted approximately as much water as Earth. Mars, more subject to impact erosion⁴, should have accreted an ocean equivalent to a layer of water ~ 10 – 100 metres deep uniformly distributed over the planet. This result is in agreement with other estimates³⁴, based on different methods, of the martian water inventory.

Note added in proof: W. K. Hartmann (personal communication and ref. 35) has considered spectrophotometric measurements of Solar System satellites thought to be captured, and concluded that asteroids scattered through the Solar System towards the end of the period of planet formation were probably $>70\%$ C type.

Received 4 August; accepted 20 November 1989.

- Anders, E. & Owen, T. *Science* **198**, 453–465 (1977).
- Chyba, C. F. *Nature* **330**, 632–635 (1987).
- McKinnon, W. B. *Nature* **338**, 465–466 (1989).
- Melosh, H. J. & Vickery, A. M. *Nature* **338**, 487–489 (1989).
- Maher, K. A. & Stevenson, D. J. *Nature* **331**, 612–614 (1988).
- Oberbeck, V. R. & Fogleman, G. *Nature* **339**, 434 (1989).
- Basaltic Volcanism Study Project *Basaltic Volcanism on the Terrestrial Planets* (Pergamon, New York, 1981).
- Neukum, G. thesis, Univ. Munich (1983).
- Wetherill, G. W. *Proc. Lunar Sci. Conf.* **8**, 1–16 (1977).
- Hartmann, W. K. in *Proc. Conf. Lunar Highlands Crust* 155–171 (Pergamon, New York, 1980).
- Grinspoon, D. H. thesis, Univ. Arizona (1988).
- Fernández, J. A. & Ip, W.-H. *Icarus* **54**, 377–387 (1983).
- Shoemaker, E. M. & Wolfe, R. F. *Proc. Lunar planet. Sci. Conf.* **15**, 780–781 (1984).
- Wilhelms, D. E. in *The Geology of the Terrestrial Planets*, NASA SP-469, 107–205 (1984).
- Wilhelms, D. E., Oberbeck, V. R. & Aggarwal, H. R. *Proc. Lunar planet. Sci. Conf.* **9**, 3735–3762 (1978).
- Melosh, H. J. *Impact Cratering* (Oxford University Press, 1989).
- Schmidt, R. M. & Housen, K. R. *Int. J. Impact Engng* **5**, 543–560 (1987).
- Shoemaker, E. M. A. *Rev. Earth planet. Sci.* **11**, 461–494 (1983).
- Haines, E. L. & Metzger, A. E. *Proc. Lunar planet. Sci. Conf.* **11**, 689–718 (1980).
- Shoemaker, E. M., Williams, J. G., Helin, E. F. & Wolfe, R. F. in *Asteroids* (ed. Gehrels, T.) 253–282 (Univ. Arizona Press, Tucson, 1979).

- Swindle, T. D., Caffee, M. W., Hohenberg, C. M. & Taylor, S. R. in *Origin of the Moon* (eds Hartmann, W. K., Phillips, R. J. & Taylor, G. J.) 331–357 (Lunar and Planetary Inst., Houston, 1986).
- Boss, A. P. & Peale, S. J. in *Origin of the Moon* (eds Hartmann, W. K., Phillips, R. J. & Taylor, G. J.) 59–101 (Lunar and Planetary Inst., Houston, 1986).
- Chyba, C. F., Jankowski, D. G. & Nicholson, P. D. *Astr. Astrophys.* **219**, L23–L26 (1989).
- Sleep, N. H., Zahnle, K. J., Kasting, J. F. & Morowitz, H. J. *Nature* **342**, 139–142 (1989).
- O'Keefe, J. D. & Ahrens, T. J. *J. geophys. Res.* **87**, 6668–6680 (1982).
- Schidowski, M. *Nature* **333**, 313–318 (1988).
- Oberbeck, V. R. & Fogleman, G. *Origins of Life* (in the press).
- Weissman, P. R. in *Geological Implications of Impact of Large Asteroids and Comets on the Earth*, *Geol. Soc. Am. Spec. Publs* No. 190 15–24 (1982).
- Delsemme, A. H. *Phil. Trans. R. Soc. A325*, 509–523 (1988).
- Mason, B. *Handbook of Elemental Abundances in Meteorites* (Gordon & Breach, New York, 1971).
- Walker, J. C. G. *Evolution of the Atmosphere* (Macmillan, New York, 1977).
- Durham, R. & Chamberlain, J. W. *Icarus* **77**, 59–66 (1989).
- Walker, J. C. G. *Icarus* **68**, 87–98 (1986).
- McKay, C. P. & Stoker, C. R. *Rev. Geophys.* **29**, 189–214 (1989).
- Hartmann, W. K. *Icarus* **71**, 57–68 (1987).

ACKNOWLEDGEMENTS. I thank C. Sagan, C. McKay, W. Hartmann, J. Kasting, D. Stevenson, P. J. Thomas, L. Brookshaw, W. McKinnon, P. Helfenstein and D. Grinspoon for discussions. I also thank P. Weissman, V. Oberbeck and H. J. Melosh for their comments on an earlier version of the manuscript. This work was supported by NASA.

Co-localization of molecules involved in antigen processing and presentation in an early endocytic compartment

Lynne E. Guagliardi*, Bruce Koppelman*, Janice S. Blum†, Michael S. Marks‡, Peter Cresswell‡ & Frances M. Brodsky*

* Department of Pharmacy, School of Pharmacy, University of California, San Francisco, California 94143, USA

† Department of Cell Biology and Physiology, Washington University School of Medicine, St Louis, Missouri 63110, USA

‡ Department of Microbiology and Immunology, Duke University Medical Center, Durham, North Carolina 27710, USA

The pathways of intracellular traffic involved in antigen processing and presentation have been defined by immunoelectron microscopy. The export pathway for class II histocompatibility molecules and the antigen import pathway meet in a peripheral endocytic compartment having all the molecular machinery believed to be required for antigen processing and presentation, including internalized surface immunoglobulins, proteolytic enzymes and invariant chains. This compartment defines a site where peptides from endocytosed antigen can bind class II molecules *en route* to the cell surface for presentation to T cells.

T-CELL receptors recognize fragments of protein antigen bound to histocompatibility molecules^{1–3}. The formation of such molecular complexes is called antigen presentation and requires the processing of antigen by denaturation and degradation⁴. Helper T cells are stimulated by class II histocompatibility molecules which generally present peptides from antigens that have entered B cells or macrophages by endocytosis⁵. Antigen uptake by B cells occurs by means of specific cell-surface immunoglobulins⁶. Cytotoxic T cells preferentially recognize class I histocompatibility molecules that bind to fragments from endogenously synthesized antigens, such as viral antigens³. The intracellular pathways involved in antigen processing and presentation have yet to be defined. But it is apparent that class I molecules bind antigenic fragments early during their biosynthesis^{7–10}, whereas class II molecules can bind peptides from exogenous antigens acquired by a different intracellular route^{5,11,12}.

During the assembly and export of class II molecules, they are transiently associated with an additional subunit called the invariant chain¹³. Transfection experiments indicate that the invariant chain is required for class II molecules to present endocytosed antigen¹⁴. It has therefore been postulated that invariant chains influence the binding of peptides to class II molecules and account for the ability of class II molecules to associate with antigenic fragments from an endocytic source¹¹. There is considerable evidence that the processing and binding of antigens to class II molecules occurs intracellularly. These events are disrupted by protease inhibitors^{15,16} and lysosomotropic drugs, and in cells deficient in endosome acidification¹⁷. Presentation is enhanced by targeting antigens to the endocytic pathway^{6,18,19}. In addition, the potential for class II molecules to encounter endocytosed antigens within cells has been demonstrated by the modification of class II molecules *en route* to the cell surface with neuraminidase conjugated to internalized transferrin²⁰, and by the co-localization of class II molecules with internalized surface immunoglobulins, as observed by using immunofluorescence microscopy²¹. To discover where these encounters occur and to examine the

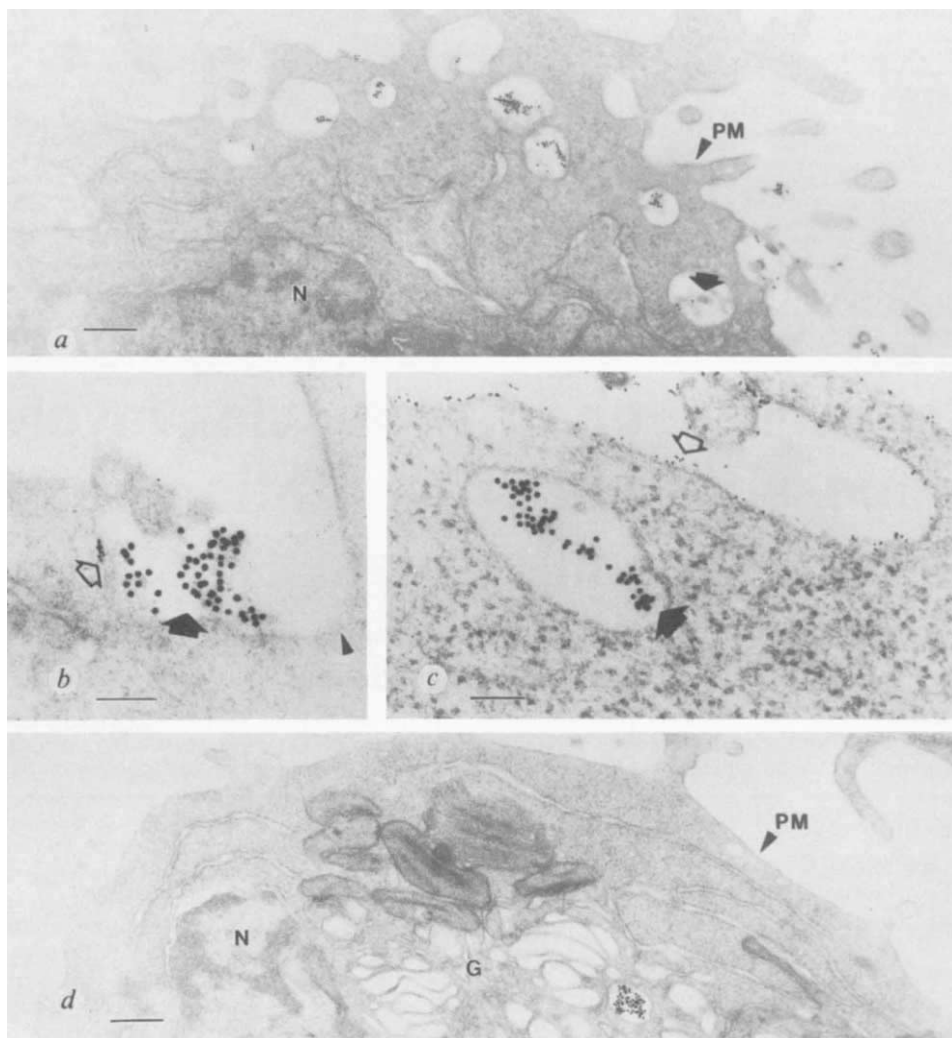
pathways involved in the processing and presentation of antigens directly, we have used electron microscopy to map the intracellular routes travelled by class II histocompatibility molecules and by endocytosed surface immunoglobulins.

We found that within 2 min of endocytosis, internalized surface immunoglobulins meet all the agents so far known to be required for antigen processing and presentation. These include both proteolytic enzymes (cathepsin B and D) and class II molecules. Invariant chains are also present in this early endocytic compartment, where the possibility of contact with endocytosed antigens would be optimized. We found that most of the intracellular class II molecules are from the export pathway. This indicates that in these cells class II molecules bind antigenic peptides primarily *en route* to the cell surface rather than after endocytosis and recycling, thereby indicating a role for invariant chains in this process.

Route of surface immunoglobulin uptake

Uptake of antigens by surface immunoglobulins leads to efficient antigen presentation by human Epstein-Barr virus-transformed B-cell lines⁶. This endocytic pathway was characterized mor-

FIG. 1 Internalization of crosslinked surface IgG by IM-9 cells. Surface IgG was crosslinked with goat anti-human immunoglobulin and rabbit anti-goat IgG conjugated to 15-nm gold beads. Cells were examined by transmission electron microscopy after incubation at 37 °C for the indicated time. PM, plasma membrane; N, nucleus; G, Golgi complex. *a*, Immunoglobulin-gold internalized for 2 min at 37 °C. Arrow shows small vesicles (30–50 nm) frequently observed within the immunoglobulin-containing compartments. Bar, 0.33 µm. *b*, Clathrin (X22 binding; 5-nm gold, clear arrow) and HLA-DR (2-nm gold, arrowhead) localized after 2 min of immunoglobulin internalization. Note the immunoglobulin (15-nm gold, dark arrow) clustered in plasma membrane invaginations. Bar, 0.12 µm. *c*, After internalization of immunoglobulin (15-nm gold, dark arrow) for 2 min, cells were labelled with wheat-germ agglutinin-gold conjugates (5-nm gold, clear arrow). Note the distribution of wheat-germ agglutinin-gold on the plasma membrane and its absence in the immunoglobulin-containing vesicle. Bar, 0.14 µm. *d*, Immunoglobulin-gold internalized for 30 min. Bar, 0.33 µm. **METHODS.** IM-9 cells²² (5×10^6) were labelled sequentially with goat anti-human immunoglobulin (Tago) (at $600 \mu\text{g ml}^{-1}$ for 60 min at 4 °C) and rabbit anti-goat IgG conjugated to 15-nm gold beads (E-Y LABS, Inc.), 1/4 dilution, 60 min at 4 °C, then incubated for 2–30 min at 37 °C. Cells were fixed (100 mM cacodylate buffer, 0.2 mM CaCl_2 , 4% sucrose, 2% paraformaldehyde; 60 min at 20 °C with 0.5% glutaraldehyde added in the last 15 min), stained with reduced osmium²⁴ and 2% aqueous uranyl acetate, and embedded in LR White Resin (medium hardness; Polysciences)²⁵. Ultra-thin sections (700 Å) were cut and mounted on unsupported 150-mesh nickel grids (Pelco). Cell sections were labelled (45 min, 4 °C) with antibodies listed in Table 1 followed by incubation (45 min, 4 °C) with the appropriate anti-immunoglobulin conjugated to 2-nm or 5-nm gold beads (1/4 dilution); (E-Y Labs, Inc.). Gold reagents were centrifuged before use to remove particulate aggregates. Rabbit antisera were amplified by binding of mouse anti-rabbit immunoglobulin (Accurate



Chemicals); ($34 \mu\text{g ml}^{-1}$, 45 min at 4 °C) before gold labelling. To detect two markers on one section, opposite sides of the grid were labelled with different primary antibodies and gold conjugates. After staining the first grid face, the labelled side was coated with a film of formvar before staining the reverse face⁶⁵. Labelled grids were viewed on a JEOL 100C transmission electron microscope at 60 kV.

TABLE 1 Antibodies to characterize intracellular compartments

Target antigen	Antibody	Form	Concentration	Ratio of vesicular-to-mitochondrial staining¶	Vesicle staining by specificity control*	
					2 nm	5 nm
HLA-DR	L243 (ref. 62)	Mouse mAb§ IgG _{2a}	10 µg ml ⁻¹	32	6	6
Invariant chain	MATILDA (ref. 26)	Rabbit antiserum	1/20	20	4	7
Cathepsin B (human)	Anti-cathepsin B*	Sheep antiserum	1/20	36	NR	6.4
Cathepsin D (human)	Anti-cathepsin D (ref. 33)†	Rabbit antiserum	1/20	104	4	7
Endocytic clathrin-coated vesicles	AP.6 (ref. 29)	Mouse mAb IgG _{2b}	10 µg ml ⁻¹	25	6	6
Clathrin	X22 (ref. 63)	Mouse mAb IgG ₁	10 µg ml ⁻¹	26	6	6
Clathrin	C77	Rabbit serum	1/20	31	4	7
DAMP (acidic vesicles)	Anti-DNP (29B5)‡	Mouse mAb IgG ₁	1 µg ml ⁻¹	21	0	5

* The Binding Site, Ltd.

† Reference gives immunization protocol. Antigen was human placental cathepsin D^{34,64}.

‡ Gift of L. Herzenberg, Stanford University.

§ Monoclonal antibody, mAb.

|| All antibodies and antisera were filtered before use (0.22 µm; Millipore).

¶ In sections labelled by each antibody, the number of gold particles in vesicles was compared with the number over an equivalent area of mitochondria present in the section, using a grid overlaying the micrograph to assess area. This gives a measure of antibody specificity, because mitochondrial labelling is not expected.

* For each antibody-gold reagent combination (and amplification antibody if used), equivalent combinations of control antibodies were tested for nonspecific labelling of immunoglobulin-containing endocytic vesicles. Percentage of immunoglobulin-containing vesicles labelled by control antibodies and gold is tabulated. A minimum of 50 vesicles per labelling combination were counted. NR, not relevant; (this size of gold particle was not used with this antibody).

phologically in the human B-lymphoblastoid cell line IM-9 (ref. 22). To mimic the binding and internalization of multivalent antigens, surface IgG was labelled at 4 °C with 15-nm gold particles. Crosslinking and internalization were induced at 37 °C, and cells were examined by electron microscopy. At 4 °C, virtually all the gold particles were randomly dispersed over the cell surface and along microvilli. After 2 min at 37 °C, the IgG-gold complexes were clustered in peripheral vesicles with electron-lucent lumina averaging 200–500 nm in diameter (Fig. 1a). Some clusters of IgG-gold were in similar-sized surface depressions surrounded by finger-like extensions of plasma membrane (Fig. 1b). This pattern of surface invagination indicated that immunoglobulin-containing vesicles observed had formed directly without passing through smaller clathrin-coated intermediates (normally ~100 nm in lymphoid cells)²³. Patches of the vesicle membranes were labelled with anti-clathrin antibodies, further indicating that the immunoglobulin-containing endocytic vesicles resulted from direct invagination, co-internalizing any coated pits which might have been forming (Fig. 1b). To determine whether the vesicles observed after 2 min were fully formed endocytic vesicles, their accessibility to wheat-germ agglutinin-gold conjugates binding to residual exposed cell-surface molecules was tested (Fig. 1c). Ninety-eight per cent of closed vesicles observed after 2 min of immunoglobulin internalization did not contain wheat-germ agglutinin, so were truly internalized endocytic vesicles, according to this criterion. After 30 min at 37 °C, internalized immunoglobulins were observed in larger vesicles (400–800 nm), some of which were present near the stacks of the Golgi lamellae (Fig. 1d). Others appeared connected to a network of tubular cisternae.

Contents of early endocytic vesicles

To identify intracellular sites for antigen processing and presentation, cells with endocytic vesicles containing internalized immunoglobulins were processed for electron microscopy. Ultra-thin sections (700 Å) were then labelled with antibodies against proteolytic enzymes, components of class II molecules and markers of intracellular traffic (Table 1). Sample preparation used minimal fixation with glutaraldehyde, reduced osmium²⁴, limited dehydration and LR White embedding medium²⁵, all selected to preserve antigenicity better than conventional fixation and embedding procedures. Antibodies bound to sections were located with 2- and 5-nm gold-antibody conjugates to contrast with the 15-nm gold marking the internalized

immunoglobulins. Nonspecific labelling with antibodies was assessed in two ways (Table 1). For each antibody, the number of gold particles labelling vesicles was compared with the number of gold particles labelling an equivalent area of mitochondria (background) in the same section. Specificity ranged from 20–104 times that of background binding. In addition, isotype-matched non-reactive monoclonal antibodies and nonimmune antisera mimicking each antibody-antiserum-gold labelling combination were assessed for their labelling of immunoglobulin-containing vesicles. Four to seven per cent of vesicles were nonspecifically labelled with single antibody combinations and 4% were nonspecifically double-labelled with two primary antibodies and two sizes of gold particles.

Specific antibody labelling showed that 2 min after induction of immunoglobulin internalization, invariant chains were present in 50% of compartments with endocytosed immunoglobulin. Clustering of labelled invariant chains (Fig. 2a, b) indicated that they could be delivered from intracellular transport vesicles such as the one shown in Fig. 2c. Because invariant chains not associated with class II molecules cannot leave the endoplasmic reticulum²⁶, those observed in these peripheral endocytic vesicles must either still be associated with class II molecules or have recently dissociated. When cell sections were labelled with a monoclonal antibody to HLA-DR (antibody L243) and with anti-invariant-chain antiserum, 41% of the endosomes containing immunoglobulins were positive for both markers. HLA-DR molecules were often separated from clusters of invariant chains in the same vesicle (Fig. 2c, d). This is consistent with a previous characterization of the specificity of L243, showing that it has a heightened affinity for mature invariant chain-negative HLA-DR (ref. 27), and indicates that mature class II molecules could segregate from invariant chain-associated molecules in these endocytic compartments (Fig. 2d). There was some staining with L243 throughout the biosynthetic pathway, indicative of either some invariant-chain dissociation in earlier compartments²⁸ or increased antibody reactivity with immature molecules as a result of fixation procedures.

Vesicles containing both internalized immunoglobulin and endogenous invariant chains were also labelled by a monoclonal antibody (AP.6) specific for the adaptin molecules of relative molecular mass 100,000, which are found exclusively in endocytic clathrin-coated vesicles²⁹ (Fig. 2e). This demonstrates an intersection of the endocytic pathway with the class II export pathway. Figure 3 summarizes the results of scoring double-

labelled sections prepared after 2 and 30 min of immunoglobulin internalization. The increase in the number of immunoglobulin-containing vesicles stained with AP.6 with time (compared with no change in clathrin staining) indicates that these adaptor proteins remain bound to endocytosed membrane even after dissociation of the clathrin coat.

To determine whether the pathway of invariant chains meets that of endocytic vesicles formed during conventional receptor-mediated endocytosis, co-localization of invariant chains with transferrin receptors was tested for. Anti-transferrin receptor monoclonal antibody (B3/25)³⁰ bound to protein A-gold complex was internalized for 2 min. Then sections were prepared and labelled with anti-invariant-chain antiserum, which stained

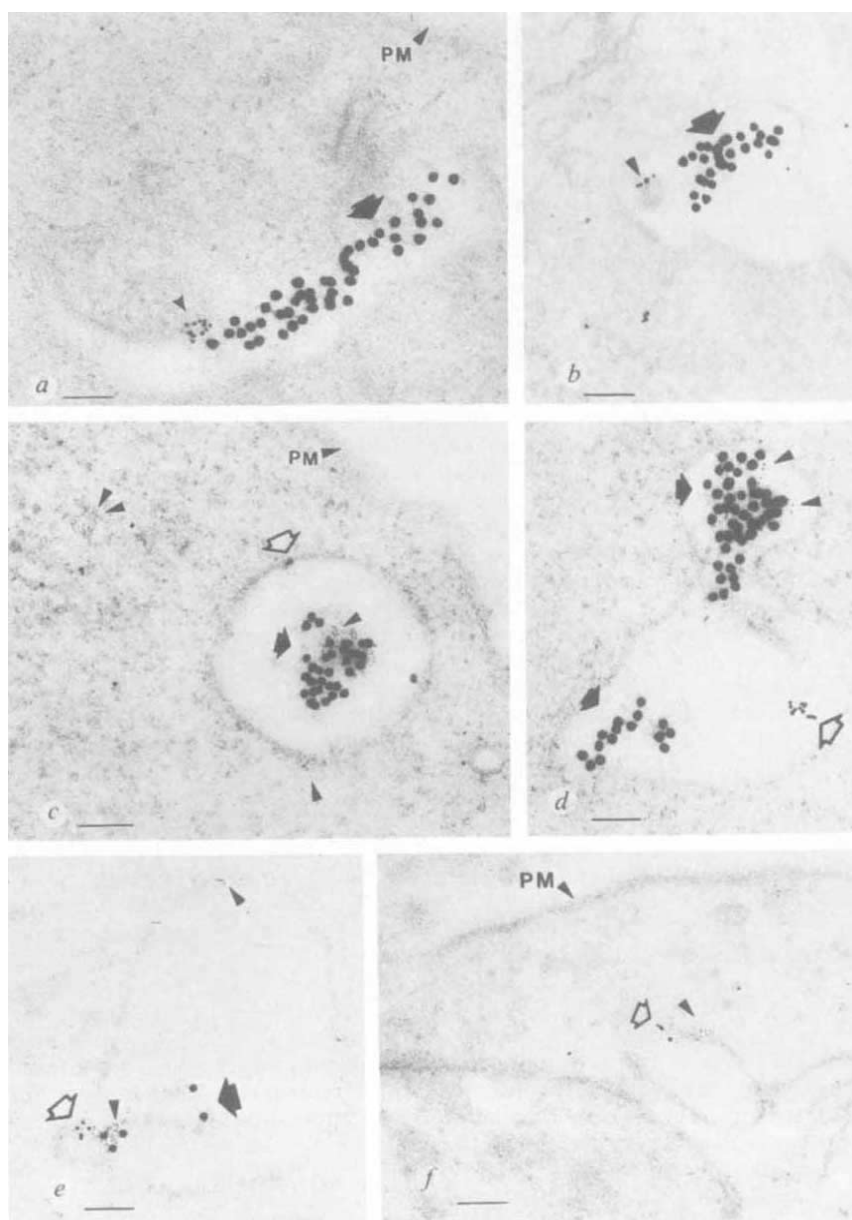
42% of endocytic vesicles with internalized transferrin receptors (Fig. 2*f*). The vesicles stained were characteristic of early endosomes formed by fusion of clathrin-coated vesicles, and typical of the well-characterized transferrin uptake pathway^{31,32}.

Most vesicles that contain immunoglobulins and invariant chains also contain HLA-DR, cathepsin B and cathepsin D (Fig. 3). Immunogold labelling of the proteases was clustered along the vesicle membrane and in adjoining tubular cisternae, again indicating that these proteases could be delivered from small transport vesicles (Fig. 4). Cathepsin B and cathepsin D isolated in endosomes are present in enzymatically active form³³⁻³⁶. Cathepsin B is active at neutral pH, whereas the optimum pH for cathepsin D is lower³⁷. Evidence that the immunoglobulin-

FIG. 2 Intracellular co-localization of HLA-DR, invariant chains and internalized immunoglobulin and transferrin receptors. Surface immunoglobulin (15-nm gold, dark arrow) was cross-linked and allowed to internalize for 2 min at 37 °C (*a-e*). Cell sections were stained as follows. *a, b*, Invariant chains localized with 5-nm gold (arrowhead). PM, plasma membrane. *a*, Bar represents 0.07 μ m. *b*, Arrowhead indicates invariant chain clustered on a smaller vesicle within immunoglobulin-containing endocytic compartment. Bar, 0.08 μ m. *c, d*, HLA-DR (5-nm gold, clear arrow) and invariant chain (2-nm gold, arrowhead). *c*, Double arrowheads show smaller vesicle with only invariant chain. Bar, 0.08 μ m. *d*, Bar represents 0.07 μ m. *e*, Endocytic adaptor protein (5-nm gold, clear arrow) and invariant chain (2-nm gold, arrowhead). Bar, 0.10 μ m. *f*, Transferrin receptor (5-nm gold, clear arrow) internalized 2 min at 37 °C, and invariant chains (2-nm gold, arrowhead). Bar, 0.10 μ m. Note that the fixation and embedding procedures used were chosen to preserve antigenicity but do not highlight membranes as well as conventional procedures. Small membranous vesicles (for example, see *c*) are identified by cytoplasmic clustering of gold label, because the outlining membrane is not preserved. Often these smaller vesicles are not electrolucent. Membranous projections into the lumen of larger vesicles were frequently observed (for example, see *e*) but not always well-delineated (for example, see *c* and *d*). What appears to be luminal staining by antibodies to membrane-bound molecules is attributed to staining of molecules associated with indistinct membranous inclusions. Clustering of gold particles is attributed to two factors. The amplification procedure used for labelling rabbit antiserum results in 2-3 gold particles per primary antibody. The rest of the clustering is interpreted as concentration in vesicles.

METHODS. Immunoglobulin internalization and section labelling were as described in the legend to Fig. 1. For internalization of transferrin receptor, cells were labelled for 60 min at 4 °C with anti-transferrin-receptor monoclonal antibody (B3/25; 40 μ g ml⁻¹) followed by protein A conjugated to 5-nm gold (E-Y Labs, Inc.)³¹. After warming to 37 °C for 2 min, cells were

prepared for electron microscopy and sections labelled with anti-invariant-chain antibody, as above.



containing endosomes acidity was provided by labelling with the dinitrophenol-derivatized amine 3-(2,4-dinitroanilino)-3'-amino-*N*-methylpropylamine (DAMP), which accumulates in intracellular compartments of low pH³⁸. After 30 min of immunoglobulin internalization, 69% of the vesicles with immunoglobulin also contained DAMP.

Quantitation of markers which co-localize with internalized immunoglobulins (Fig. 3) indicates that from 2–30 min after endocytosis, immunoglobulins are present in an endocytic compartment with proteases required for antigen processing and class II molecules required for antigen presentation. The level of co-localization observed is a minimum estimate, because the efficiency of antibody binding to proteins treated for electron microscopy is reduced to a variable degree, depending on the nature of individual antibody-antigen recognition sites.

Intracellular origins

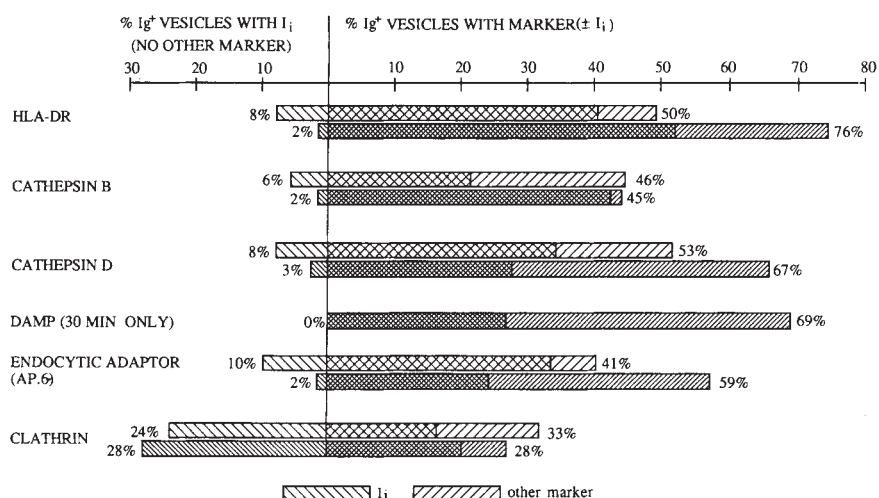
In the analysis described here of the intracellular distribution of invariant chains using antiserum and colloidal gold, there was no significant labelling of the cell surface. Pulse-chase studies indicate that invariant chains are localized to the biosynthetic pathway¹³; but for some cells a small proportion can be detected on the cell surface³⁹ (N. Koch, personal communication). To measure the surface expression of invariant chains on IM-9 cells, they were labelled with anti-invariant-chain antiserum and analysed by flow cytometry (Fig. 5a). No significant staining was detected, indicating that for IM-9 cells the invariant chains in immunoglobulin-containing vesicles were derived from an internal pool, and not from endocytosis of surface invariant chains. It is possible that the antiserum used may not react with a surface form of invariant chain. However, if this is the case, then the intracellular labelling with antibody can only be derived from internal invariant chain. Flow cytometric analysis of surface labelling with anti-cathepsin-B and anti-cathepsin-D antisera also revealed no significant binding over that of background (data not shown). These results correlate with the lack of observed surface labelling with colloidal gold when these antisera were bound to cell sections. Previous studies of the intracel-

lular distribution of cathepsin D indicate that its presence in endosomes is probably derived from an intracellular³³ rather than endocytic route. Small vesicles containing cathepsin D have been observed in the Golgi region and are believed to transport cathepsin D to endosomes^{33,34}. Although no evidence for surface cathepsin D has been found⁴⁰, enzymatic assays have revealed cathepsin B activity in plasma membrane-enriched cell fractions^{16,41}, and cathepsin B has been detected with antibody on the surface of tumour cells⁴². In the former analysis, the detected cathepsin B activity could have been derived from endosomes co-purified with the plasma membrane fraction^{35,36}. In the latter case, it has been proposed that surface expression of cathepsin B was related to the metastatic state of the tumour. Our results indicate that surface expression of proteases could vary with cell type, similar to the variation observed for surface expression of invariant chains.

Because IM-9 cells have high levels of surface HLA-DR, it is possible that mature class II molecules could enter immunoglobulin-containing endosomes by constitutive endocytosis from the plasma membrane or by co-internalization with crosslinked immunoglobulins, as well as by the class II export pathway. To study the spontaneous internalization of HLA-DR, cells were labelled with iodinated L243 Fab fragments at 4 °C. After warming for various times to allow endocytosis, cells were chilled and the remaining surface Fab fragments stripped off by an acid wash. Radioactivity that remained associated with the cells indicated that only a small percentage (1.5–3.5%) of surface HLA-DR was spontaneously internalized by these cells (Fig. 5b). By contrast, a similar experiment with iodinated anti-transferrin-receptor monoclonal antibody³⁰ showed that high levels of transferrin receptor were internalized (peaking at 51%) and that receptors were recycled later on. Crosslinking and endocytosis of surface immunoglobulins induced by colloidal gold only increased the internalization of HLA-DR molecules to a maximum of 3.7% (Fig. 5b), showing that most were unaffected. Thus, very little HLA-DR is found in endocytic vesicles as a result of 'bystander' internalization along with crosslinked immunoglobulin. This indicates that most

FIG. 3 Frequency of internalized immunoglobulin co-localization with molecules involved in antigen processing and presentation. Cell sections were prepared after 2 min and 30 min of immunoglobulin internalization and double-labelled for the presence of invariant chain (I_i) in addition to the other markers listed in Table 1, as described in Fig. 1. The bars to the right of the central line show the percentage of immunoglobulin-containing (Ig⁺) vesicles labelled for the marker indicated and the proportion (cross-hatched) labelled with both invariant chain and the marker. To the left, the bars show percentage of Ig⁺ vesicles with invariant chain and no additional marker for each double-labelling experiment. Results from scoring Ig⁺ vesicles after 2 min of internalization are given in the upper bar in each pair; the lower bars are results after 30 min of internalization. DAMP³⁸ labels acidic vesicles and was detected with anti-DNP monoclonal antibody (29B5). It was only localized in sections of cells internalizing immunoglobulin-gold for 30 min because its sequestration requires a 30-min incubation. Clathrin was localized using the C77 antiserum in this experiment.

METHODS. A minimum of 50 immunoglobulin-containing vesicles were counted for each labelled section. These were randomly selected at the electron microscope and represent cumulative results from ~50 cells, for each labelling combination. Data tabulated was collected from two separate



internalization experiments which gave similar results. Two preliminary experiments (data not included) also correlated with these statistics. For levels of background labelling, see Table 1. To label cells with DAMP, surface immunoglobulin was first labelled at 4 °C as in Fig. 1. Cells were then incubated for 30 min at 37 °C with medium containing 50 μM DAMP (Molecular Probes, Inc.) before fixation and sectioning. C77 anti-clathrin antisera was made against clathrin baskets (polymerized without adaptin molecules) according to the method of Bloom and Puszkis⁶⁶. This antiserum recognizes clathrin heavy chains on immunoblots.

of the intracellular class II molecules detected by L243 represent molecules that have recently dissociated from invariant chains on their way to the cell surface. Similar low levels of constitutive HLA-DR internalization have been demonstrated for other human B-cell lines⁴³ by using a ligand-independent assay that measures inaccessibility to neuraminidase⁴⁴.

Discussion

These studies demonstrate that the antigen import pathway and the export pathway for class II histocompatibility molecules meet in a peripheral endocytic compartment. This compartment has all the molecular machinery that is known to be required for antigen processing and presentation, including proteolytic enzymes and invariant chains. It defines an intracellular site where peptides from endocytosed antigen could bind class II molecules *en route* to the cell surface.

The antigen import pathway has been analysed using gold particles to crosslink surface immunoglobulins and induce internalization. Crosslinking of surface immunoglobulins triggers intracellular signalling comparable to that produced by the binding of multivalent antigens to B cells⁴⁵. Endocytosis of crosslinked surface immunoglobulins can therefore be considered to mimic the uptake of multivalent antigens. The uptake pathway defined here could also be relevant to the endocytic fate of monovalent antigens. Watts *et al.*⁴⁶ showed that tetanus toxoid coupled to colloidal gold is endocytosed by surface immunoglobulins in clathrin-coated vesicles, whereas the data presented here show that more-extensively crosslinked surface immunoglobulins are internalized nonselectively. Tran *et al.*⁴⁷ demonstrated, however, that ligands taken up by either selective or nonselective endocytosis are delivered to the same endosomes. Thus the internalization mechanisms for multivalent and monovalent antigens could differ according to the degrees of immunoglobulin crosslinking, but the subsequent endocytic pathways are probably similar. This is demonstrated further by the comparison described above of the uptake pathway of transferrin receptors with that of crosslinked surface immunoglobulins. Invariant chains, marking the class II export pathway, intersect the early endocytic compartments containing transferrin receptors to the same extent that they co-localize with internalized crosslinked surface immunoglobulins.

Our results emphasize the importance of the class II export pathway in antigen presentation. B-lymphoblastoid cell lines that internalize antigens by means of surface immunoglobulins, present antigens very efficiently to T cells⁶. We have demonstrated here and previously⁴³ that these cells internalize very little of their surface class II molecules. In this regard, these cells differ from those of the mouse B-cell lymphoma studied by Harding and Unanue⁴⁸, which internalize up to 35% of

surface class II molecules. Thus the endocytosis-peptide exchange-recycling mechanism proposed for antigen presentation in the mouse B-cell lymphoma is unlikely to operate in the cells used in this study, although such a mechanism may be relevant for cells that internalize their class II molecules. For these human B-lymphoblastoid cell lines, peptide binding by class II molecules is most likely to occur as they travel to the cell surface. The presence of invariant chains in the endocytic compartments where class II molecules are localized is additional evidence that these class II molecules are derived from the export pathway¹³. Biosynthetic studies indicate that export of class II molecules is relatively slow⁴⁹⁻⁵¹. In both human and mouse B-cell lines, there is a large intracellular pool of class II molecules associated with invariant chains, detectable 2-4 h after pulse labelling^{48,51}. This explains why antigen presentation mediated by these molecules is insensitive to inhibitors of protein synthesis^{12,48}. For mouse macrophages, in which intracellular pools of class II molecules are reduced by adherence, antigen presentation is sensitive to cycloheximide⁴⁸.

Finding that the invariant chains accompany class II molecules to a peripheral endocytic compartment is morphological evidence for their role in antigen presentation. These results are consistent with either of two models for the function of invariant chains¹¹. The first model proposes that invariant chains occupy or stabilize the peptide-binding site during the assembly of class II molecules, so it does not need to bind peptides to exit from the endoplasmic reticulum, as seems to be required for class I molecules. Once the invariant chain-class II complex reaches peripheral endocytic compartments, it encounters proteases cathepsin B and D implicated in cleavage of the invariant chain, correlating with its dissociation^{52,53}. Because cathepsins are also involved in antigen processing^{15,16}, the immunoglobulin-containing endocytic compartments characterized here provide a potential location for dissociation of invariant chains in the presence of antigenic peptides, which can then bind to the revealed site. Gradual endosome acidification⁵⁴ could delay the action of cathepsin D and thereby slow down invariant-chain and antigen processing. This would explain why the processing events take 30-60 min^{55,56}, even though the appropriate molecules can co-localize within 2 min of endocytosis.

A second proposed role for invariant chains which is consistent with the data described here is to target class II molecules to endocytic compartments¹¹. Here they could exchange peptides acquired from endogenous proteins for those generated from endocytosed proteins. In this case, class II molecules could bind peptide in the endoplasmic reticulum but exchange peptides more readily than class I molecules, particularly in the low pH environment of the endosome. This hypothesis is supported by the observations that purified class II molecules bind peptide

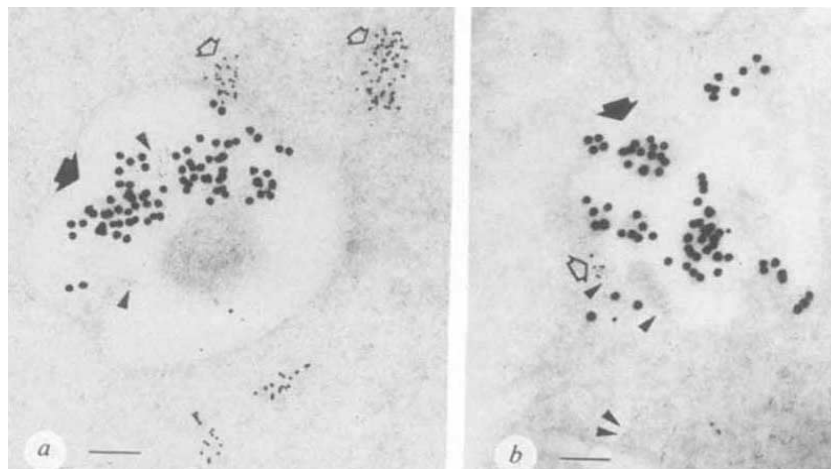
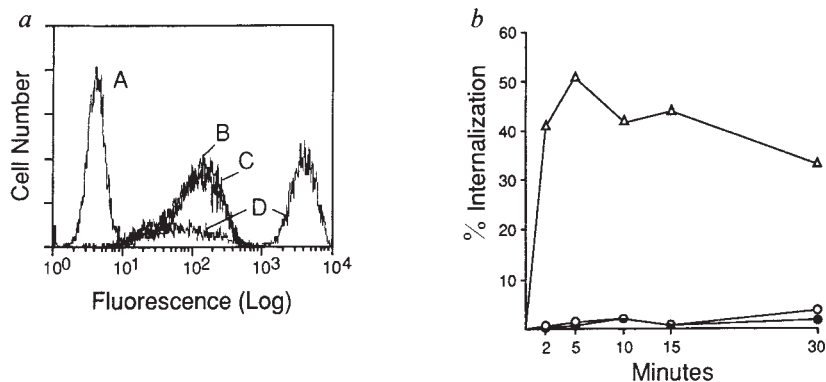


FIG. 4 Invariant chains and internalized immunoglobulins co-localize in protease-containing compartments. *a*, Cathepsin B (5-nm gold, clear arrow) and invariant chains (2-nm gold, arrowhead) localized after 2 min of immunoglobulin (15-nm gold, dark arrow) internalization. Bar, 0.10 μ m. *b*, Cathepsin D (5-nm gold, clear arrow) and invariant chains (2-nm gold, arrowhead) localized after 30 min of immunoglobulin (15-nm dark arrow) internalization. This is also typical of cathepsin D distribution in compartments seen after 2 min immunoglobulin uptake. Note the small vesicle containing only invariant chains (double arrowheads). Bar, 0.07 μ m.

FIG. 5 Origin of invariant chains and class II molecules in endocytic vesicles. *a*, Surface staining for invariant chain. FACSCAN flow cytometer analysis of IM-9 cells labelled with anti-invariant-chain (B) or pre-immune antiserum (C), anti-HLA-DR monoclonal antibody (L243) and rabbit anti-mouse immunoglobulin antiserum (D), or anti-rabbit IgG-FITC alone (A). *b*, Internalization of HLA-DR and transferrin receptor. Cells were incubated with 125 I-labelled L243 Fab (anti-HLA-DR) (●), 125 I-labelled B3/25 IgG anti-transferrin receptor (△) or with reagents to crosslink surface IgG in addition to 125 I-labelled L243 Fab (○). Iodinated antibodies were internalized at 37 °C and radioactivity remaining on the surface was stripped off by acid wash⁶⁷ at indicated time points.



METHODS. *a*, IM-9 cells (10^6) were incubated with anti-invariant-chain and preimmune antisera (1/100, 45 min at 4 °C) followed by fluorescein isothiocyanate-conjugated (FITC) anti-rabbit immunoglobulin (1/2,000, 45 min at 4 °C). For HLA-DR labelling, cells (10^6) were incubated with L243 IgG (at 10 μ g ml⁻¹ for 45 min at 4 °C) followed by rabbit anti-mouse immunoglobulin serum (1/100, 45 min at 4 °C), followed by FITC-conjugated anti-rabbit IgG. *b*, L243 Fab fragments produced using pepsin⁶⁸, and B3/25 IgG were iodinated using Iodobeads (Pierce Chemicals), 25 μ g protein with 1 mCi 125 I-labelled NaI for 5 min. IM-9 cells (10^7) were incubated with 5×10^7 c.p.m. iodinated antibody for 1 h at 4 °C, washed and aliquots incubated at 37 °C for 2–30 min. Cells were washed at each time point and exposed to 150 μ l HCl (20 mM), 150 mM NaCl, pH 1.7 (3 min)⁶⁷ or PBS,

pH 7.4. The percentage of surface antibody internalized was calculated from the number of cell-associated counts after acid stripping, divided by the number of counts associated with unstripped cells. Background counts remaining associated with acid-stripped cells incubated at 4 °C were subtracted before calculating the percentage internalization. Samples analysed after acid-stripping showed a loss of 10% of the labelled cells. To measure internalization in the presence of crosslinked immunoglobulin, surface immunoglobulin was labelled at 4 °C with goat anti-immunoglobulin followed by 15-nm gold-anti-goat immunoglobulin in addition to iodinated antibody. The internalization and acid-stripping steps were performed as above.

more readily than do purified class I molecules^{57,58}. This mechanism for the binding of peptides by class II molecules explains the apparent peptide exchange in cells which have some degree of internalization of class II molecules⁵⁹. Both models for the function of invariant chains account for the ability of class II molecules to present peptides from some endogenously synthesized viral proteins^{60,61} as well as endocytosed antigens. In the first model, these peptides compete

with invariant chains in the endoplasmic reticulum, and in the second model class II molecules bind peptides in the endoplasmic reticulum that are resistant to exchange. Alternatively, there could be an intracellular route by which peptides from endogenously synthesized antigens can access endosomes. To determine whether invariant chains directly influence peptide binding or are responsible for directing class II molecules to endocytic compartments, or both, will require further study. □

Received 1 September; accepted 22 November 1989.

1. Bjorkman, P. J. *et al.* *Nature* **329**, 506–512 (1987).
2. Watts, T. H. & McConnell, H. M. *A. Rev. Immun.* **5**, 461–475 (1987).
3. Townsend, A. & Bodmer, H. *A. Rev. Immun.* **7**, 601–624 (1989).
4. Unanue, E. R. *A. Rev. Immun.* **2**, 395–428 (1984).
5. Germain, R. N. *Nature* **322**, 657–682 (1986).
6. Lanzavecchia, A. *Nature* **314**, 537–539 (1985).
7. Townsend, A. *et al.* *Nature* **340**, 443–448 (1989).
8. Parham, P. *Nature* **340**, 426–428 (1989).
9. Nuchtern, J. G., Bonifacio, J. S., Biddison, W. F. & Klausner, R. D. *Nature* **339**, 223–226 (1989).
10. Yewdell, J. W. & Bennink, J. R. *Science* **244**, 1072–1075 (1989).
11. Long, E. O. *Immun. Today* **7**, 232–234 (1989).
12. Braciale, T. J. *et al.* *Immunol. Rev.* **98**, 95–114 (1987).
13. Cresswell, P., Blum, J. S., Kelnar, D. N. & Marks, M. S. *CRC Crit. Rev. Immun.* **7**, 31–53 (1987).
14. Stockinger, B. *et al.* *Cell* **56**, 683–689 (1989).
15. Berzofsky, J. A., Brett, S. J., Streicher, H. Z. & Takahashi, H. *Immunol. Rev.* **106**, 5–31 (1988).
16. Chain, B. M., Kay, P. M. & Shaw, M.-A. *Immunol. Rev.* **106**, 33–58 (1988).
17. McCoy, K. L. *et al.* *Immun.* **143**, 29–38 (1989).
18. Pierce, S. K., Morris, J. F., Grusby, M. J., Kaumaya, P., Van Buskirk, A., Srinivasan, M., Crump, B. & Smolenski, L. A. *Immunol. Rev.* **106**, 149–180 (1988).
19. Snider, D. P. & Segal, D. M. *J. Immun.* **143**, 59–65 (1989).
20. Cresswell, P. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8188–8192 (1985).
21. Pletscher, M. & Pernis, B. *Eur. J. Immun.* **13**, 581–584 (1983).
22. Fahey, J. L., Buell, D. N. & Sox, H. C. *Ann. N.Y. Acad. Sci.* **190**, 221–234 (1971).
23. Salisbury, J. L., Condeelis, J. S. & Satir, P. *J. Cell Biol.* **87**, 132–141 (1980).
24. Coggeshall, R. E. *Anat. Rec.* **194**, 201–204 (1979).
25. Newman, G. R., Jasani, B. & Williams, E. D. *Histochem. J.* **15**, 543–555 (1983).
26. Cresswell, P., Blum, J. S., Marks, M. S. & Roche, P. A. *Cold Spring Harb. Symp. quant. Biol.* **54**, 309–318 (1989).
27. Shackelford, D. A., Lampson, L. A. & Strominger, J. L. *J. Immun.* **127**, 1403–1410 (1981).
28. Nguyen, Q. V. & Humphreys, R. E. *J. Biol. Chem.* **164**, 1631–1637 (1989).
29. Chin, D. J., Straubinger, R. M., Acton, S., Nätthke, I. & Brodsky, F. M. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9289–9293 (1989).
30. Omary, M. B., Trowbridge, I. S. & Minowada, J. *Nature* **286**, 888–891 (1980).
31. Iacopetta, B. J., Rothenberger, S. & Kühn, L. C. *Cell* **54**, 485–489 (1988).
32. Dautry-Varsat, A., Ciechanover, A. & Lodish, H. F. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2258–2262 (1983).
33. Blum, J. S. *et al.* *Cold Spring Harb. Symp. quant. Biol.* **54**, 287–292 (1989).
34. Diment, S., Leech, M. S. & Stahl, P. D. *J. Biol. Chem.* **263**, 6901–6907 (1988).
35. Wiley, S. H., Van Nostrand, W., McKinley, D. N. & Cunningham, D. D. *J. Biol. Chem.* **260**, 5290–5295 (1985).
36. Roederer, M., Bowser, R. & Murphy, R. F. *J. Cell. Physiol.* **131**, 200–209 (1987).
37. Barrett, A. J. & McDonald, J. K. *Mammalian Proteases* Vol. 1 (Academic, New York, 1980).

38. Anderson, R. G., Falck, J. R., Goldstein, J. L. & Brown, M. S. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4838–4842 (1984).
39. Knapp, W. *et al.* *Immun. Today* **10**, 253–258 (1989).
40. Rodman, J. S., Levy, M. A., Diment, S. & Stahl, P. D. *J. Leukocyte Biol.* (in the press).
41. Sloane, B. F., Rozhin, J., Hatfield, J. S., Crissman, J. D. & Honn, K. V. *Exp. cell. Biol.* **55**, 209–224 (1987).
42. Graf, M., Baici, A. & Sträubli, P. *Lab. Invest.* **45**, 587–596 (1981).
43. Davis, J. E. & Cresswell, P. *J. Immun.* (in the press).
44. Krangel, M. S. *J. exp. Med.* **165**, 1141–1159 (1987).
45. Myers, C. D., Kriz, M. K., Sullivan, T. J. & Vitetta, E. S. *J. Immun.* **138**, 1705–1711 (1987).
46. Watts, C., West, M. A., Reid, P. A. & Davidson, H. W. *Cold Spring Harb. Symp. quant. Biol.* **54**, 345–352 (1989).
47. Tran, D., Carpentier, J. L., Sawano, F., Gorden, P. & Orci, L. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7957–7961 (1987).
48. Harding, C. V. & Unanue, E. R. *J. Immun.* **142**, 12–19 (1989).
49. Ploegh, H. & Fuhrmann, U. in *Cell Biology of the Major Histocompatibility Complex* (eds Pernis, B. & Vogel, H. J.) 133–151 (Academic, Orlando, 1985).
50. Nowell, J. & Quaranta, V. *J. exp. Med.* **162**, 1371–1376 (1985).
51. Machamer, C. E. & Cresswell, P. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1287–1291 (1984).
52. Blum, J. S. & Cresswell, P. *Proc. natn. Acad. Sci. U.S.A.* **85**, 3975–3979 (1988).
53. Nguyen, Q. V., Knapp, W. & Humphreys, R. E. *Hum. Immun.* **24**, 153–163 (1989).
54. Mellman, I., Fuchs, R. & Helenius, A. *A. Rev. Biochem.* **55**, 663–700 (1986).
55. Ziegler, H. K. & Unanue, E. R. *Proc. natn. Acad. Sci. U.S.A.* **79**, 175–178 (1982).
56. Davidson, H. W. & Watts, C. *J. Cell Biol.* **109**, 85–92 (1989).
57. Buus, S., Sette, A., Colon, S. M., Jenis, D. M. & Grey, H. M. *Cell* **47**, 1071–1077 (1986).
58. Chen, B. P. & Parham, P. *Nature* **337**, 743–745 (1989).
59. Harding, C. V., Roof, R. W. & Unanue, E. R. *Proc. natn. Acad. Sci. U.S.A.* **86**, 4230–4234 (1989).
60. Jin, Y., Shih, J. W.-K. & Berkower, I. *J. exp. Med.* **168**, 293–306 (1988).
61. Jacobson, S., Sekaly, R. P., Jacobson, C. L., McFarland, H. F. & Long, E. O. *J. Virol.* **63**, 1756–1769 (1989).
62. Lampson, L. A. & Levy, R. *J. Immun.* **125**, 293–299 (1980).
63. Brodsky, F. M. *J. Cell Biol.* **101**, 2047–2054 (1985).
64. Takahashi, T. & Tang, J. *Meth. Enzym.* **80**, 565–581 (1981).
65. Descheppe, C. F., Crumrine, D. A. & Ganong, W. F. *Endocrinology* **119**, 36–43 (1986).
66. Bloom, W. S. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **77**, 5520–5524 (1980).
67. Watts, C. & Davidson, H. W. *EMBO J.* **7**, 1937–1945 (1988).
68. Parham, P. in *Handbook of Experimental Immunology* Vol. 1 (eds Weir, D. M., Herzenberg, L. A., Blackwell, C. & Herzenberg, L. A.) 14.1–14.23 (Blackwell Scientific Publications, Oxford, 1988).

ACKNOWLEDGEMENTS. We thank D. Crumrine for technical assistance, D. Wong for computer graphics, D. Crommie for help with the FACS analysis, I. Trowbridge and L. Herzenberg for antibodies, and E. Belenson for manuscript preparation. This work was supported by the NIH (F.B. and P.C.), the NSF (F.B.), the Pew Charitable Trusts (F.B.) and postdoctoral awards from the California Chapter of the American Heart Association (L.G.) and the Leukemia Society of America (J.B.).

Structural basis of the allosteric behaviour of phosphofructokinase

Tilman Schirmer* & Philip R. Evans

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Comparison between the crystal structures of low- and high-affinity forms of phosphofructokinase shows a close coupling between the change of quaternary structure and local changes triggered by binding of the allosteric effectors. These concerted changes link all the substrate and effector sites in the tetramer, and explain the change of affinity for the cooperative substrate.

SINCE the pioneering work of Monod and Koshland and their colleagues^{1,2}, allosteric regulation of enzyme activity has been attributed to the switching between two (or more) conformations of the enzyme, but except for haemoglobin³, the molecular mechanism for this switch has begun to be understood only recently: now the structures of aspartate transcarbamylase⁴ and glycogen phosphorylase⁵ explain at least partially how they work. Here we show how binding of the allosteric inhibitor to phosphofructokinase switches its conformation to a form with a low affinity for substrate.

Phosphofructokinase (PFK) from *Escherichia coli* was one of the first enzymes to which the Monod model was applied⁶. It is the major regulator of flux through the glycolytic pathway, and shows sigmoidal kinetics with respect to one substrate, fructose-6-phosphate (Fru6P), but not to the other substrate ATP; allosteric activation by ADP or GDP, and allosteric inhibition by phosphoenolpyruvate (PEP). PFK from *Bacillus stearothermophilus* behaves similarly, except that the kinetics with respect to Fru6P are sigmoidal only in the presence of PEP⁷; that is, the low-affinity T state is only present with PEP. Fifty-four per cent of the amino acids of these two PFKs are identical, and they have very similar structures. Structures of several crystal forms of both these bacterial PFKs have been determined, in attempts to characterize the forms with high and low affinity for Fru6P, which we label as R and T following Monod. The structure of the high-affinity active form (R state) with bound substrates and activator has been well characterized for the enzymes from both species⁸⁻¹⁰, and provides much information about the mechanism of catalysis. Surprisingly, crystals of *E. coli* PFK grown in the absence of ligands had an R structure¹¹. Here we describe the determination of the inhibited T structure of *B. stearothermophilus* PFK to 2.5 Å resolution, starting from an earlier structure at low resolution¹². Comparison with the R structure shows that the quaternary change first seen at low resolution is well approximated by a relative rotation of rigid dimers in the tetramer. This rotation is coupled to the restructuring of a polypeptide loop which interchanges the positions of a vital arginine and a glutamate in the substrate-binding site, and to the insertion (or removal) of a layer of water molecules between pairs of subunits.

Structure of the enzyme

B. stearothermophilus PFK is a tetramer of identical subunits with 319 amino acids each. The structure is outlined in Fig. 1,

and the R structure of the homologous *E. coli* PFK is described in ref. 10. Each subunit consists of two domains, a larger one which binds the substrate ATP, and a smaller one which binds Fru6P (Fig. 1c): both domains are 3-layered $\alpha\beta\alpha$ sandwich structures. The subunits pack together with a large interface to form dimers (subunits AB or CD in Fig. 1), and the dimers pack

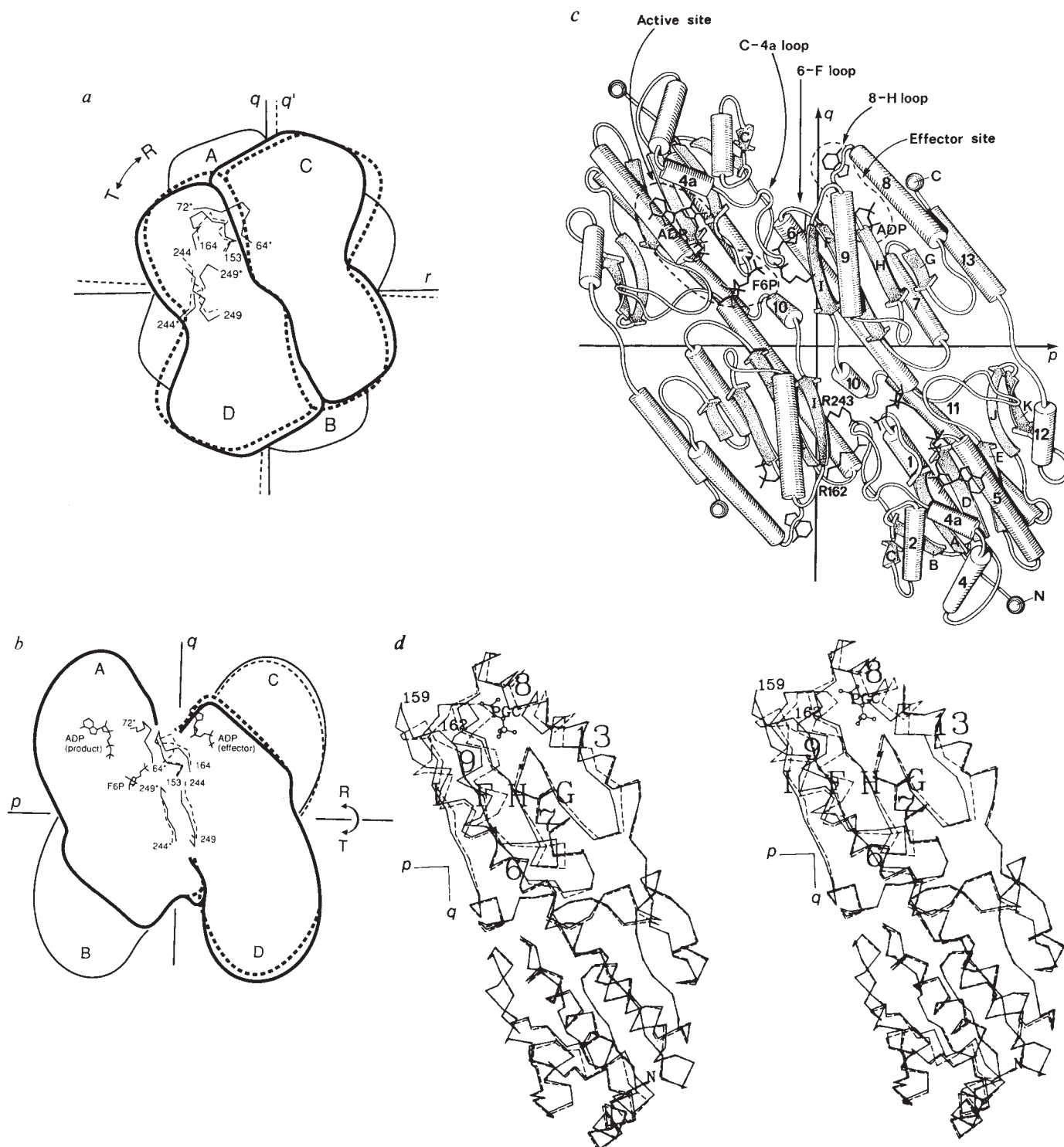
FIG. 1 General view of PFK and its quaternary and tertiary structure changes on the allosteric transition from R to T. a, Schematic sketch of the R-state (dashed outline) and T-state (solid outline) PFK tetramer, viewed along p . The three mutual perpendicular 2-fold symmetry axes are labelled p , q and r . The subunits are labelled A–D. Both structures have been superimposed after fitting the respective AB dimers. Part of the C_α -tracing (R state, dashed; T state, solid lines) in the dimer–dimer interface is depicted, residue labels with asterisk belong to subunit A, all others to subunit D. Dimer CD performs a relative rotation of $\sim 7^\circ$, approximately around the p axis. Note the backbone rearrangement of residues 156–162. b, Same as in a but viewed along r and additional representation of the ligands of the R structure. Note the closure of the gap between the backbones 244*–249* and symmetry-related partner 244–249 in the T state. c, R-state structure, subunits A and D, related by r dyad, view as in b. Helices are represented as cylinders numbered 1–13, β -sheet strands as arrows labelled A–K. Note the tenuous (interdimer) interface between the subunits. In the active site the substrate Fru6P (F6P) and the product ADP are shown, in the effector site the activator ADP. The side chains of Arg 162 (R 162) and Arg 243 (R 243) bind the 6-phosphate group of Fru6P across the subunit interface from subunit A–D, and vice versa. d, Superposition of R-state (dashed lines) and T-state (solid lines) subunit (see Table 2), view as subunit D in b and c. The subunit consists of a small (top) and a large (bottom) domain. Note shear motion of helices 8 and 9 on top of the β -sheet of the small domain: these helices were omitted from the superposition. The inhibitor 2-phosphoglycolate (PGC) is shown in the effector site.

METHODS. A plasmid based on pBR322 expressing *B. stearothermophilus* phosphofructokinase^{7,13} was transformed into DF1020 (*pfk*[−] *E. coli* strain)¹⁴. The protein was prepared essentially as described¹⁰, eluting from the Amicon blue-A column with 5 mM ATP, 10 mM MgCl₂ and 0.5 M NaCl. Crystals (typical dimensions 0.7 × 0.15 × 0.15 mm) were grown from 1.6 to 1.8 M potassium tartrate (pH 7.8), 10 mM 2-phosphoglycolate (PGC), 0.5 mM EDTA, 0.5 mM dithiothreitol (DTT), 9–6 mg ml^{−1} protein¹². Diffraction data were collected with a FAST TV-diffractometer with crystal-to-detector distance 70 mm, detector angle 18°, using Ni filtered radiation from a rotating anode X-ray generator. The crystals belong to space group P2₁2₁2₁ (refined cell constants $a=131.85$, $b=115.29$, $c=96.06$ Å). Data from three crystals were collected at a speed of 2.4 degrees per hour using the program MADNES. The data were split into 41 batches of rotation width 3° and scaled together. 86,496 measurements were reduced to 43,193 unique reflections to 2.5 Å resolution (85% complete) with $R_{\text{merge}}=0.11$. The model from the previous 7 Å analysis of this crystal form¹² showed good agreement with the data, $R=0.35$ for data to 3-Å resolution. Four cycles of restrained refinement using the programs DERIV (ref. 15) and PROLSQ (ref. 16) dropped the R factor to 0.28. This model was used only to derive a starting set of phases, a molecular envelope, and the positions of the local 222 axes. Seven cycles of fourfold averaging¹⁷ of weighted electron density maps¹⁸ at 3.0-Å resolution improved the averaging R factor from 0.27 to 0.19, then the resolution was increased stepwise to 2.6-Å resolution, giving a final $R_{\text{average}}=0.19$. In most parts, the averaged map was easily interpretable. Density was visible for the inhibitor in all four subunits. The remodelled structure was refined further, with the 222 symmetry of the model maintained at first. The current model (including 185 water molecules, with a r.m.s.d. of the bond lengths from target values of 0.010 Å) gives an R factor of 0.183 for all measured reflections to 2.5-Å resolution, including a bulk solvent contribution to the F_{calc} (ref. 10). For comparison with the new T-state structure, the R-state structure⁹ incorporating Fru6P and ADP/Mg²⁺ was refined further, to an R factor of 0.172 for all data to 2.4 Å resolution.

* Present address: Biozentrum der Universität Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland.

together with a smaller interface to form the tetramer. The tetramer exhibits 222 symmetry: the three molecular dyad axes will be referred to as p , q and r (Fig. 1). The large interface within the dimer (around the p dyad) includes the allosteric effector site, which binds both activator and inhibitor. The small interface between dimers (around the r dyad) is bridged in the R state by the side chains of Arg162 and Arg243, which bind the 6-phosphate group of a Fru6P in the other dimer. This position of the cooperative substrate Fru6P in a subunit interface⁸ suggested that the dimer-dimer interface would be changed in the transition between the T and R state, and indeed such a change was detected in the low-resolution structure of the T state¹².

The enzyme was isolated from cultures of *E. coli* containing a plasmid expressing *B. steartophilus* PFK at a high level (see legend to Fig. 1). Crystals of the T state were grown as before¹² in the absence of activating ligands and in the presence of the nonphysiological inhibitor 2-phosphoglycolate, a stable analogue of PEP. The improved quality of the crystals, which now diffract to beyond 2.5 Å resolution, may be due to purer protein obtained with the high levels of expression from the plasmid, compared to the original isolation from the bacterium. The structure was determined by a combination of molecular averaging and refinement, starting from the previously determined model: details are given in the legend to Fig. 1. This crystal form contains the entire tetramer in the asymmetric unit,



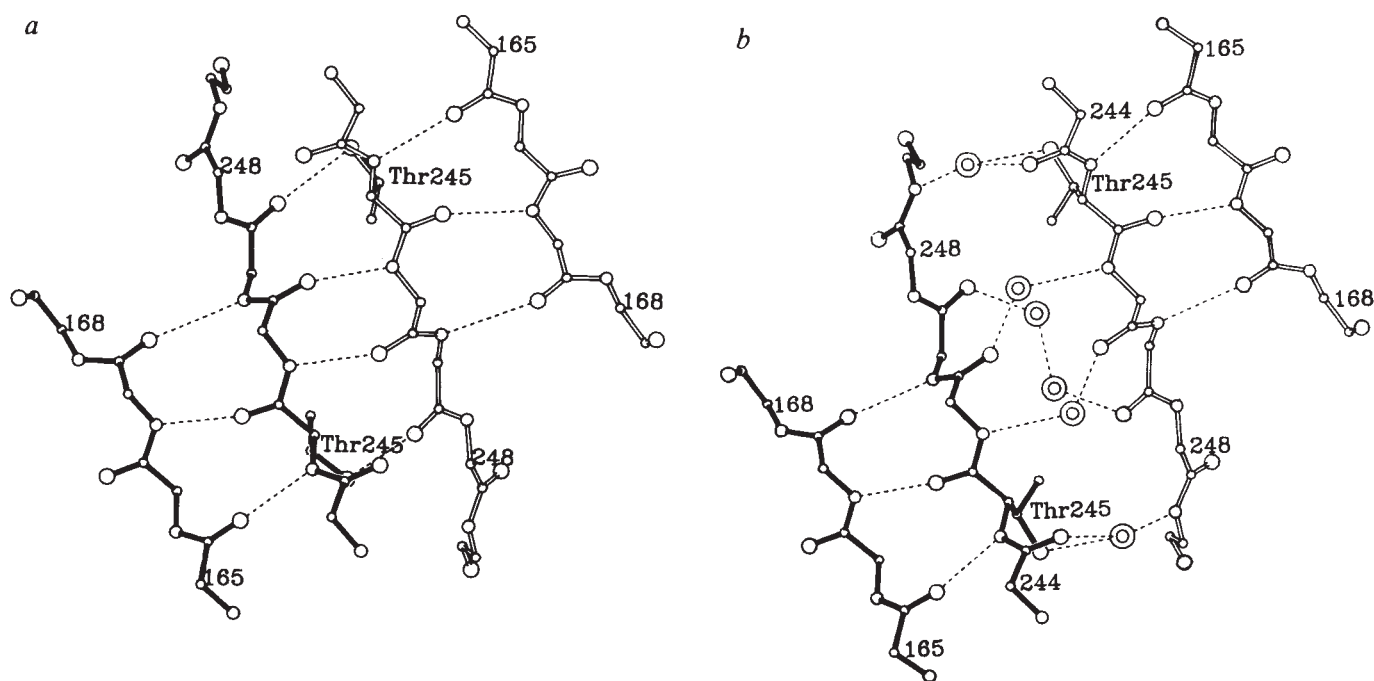


FIG. 2 Interactions between dimers across the *r* axis, showing the passive change between R and T state, view as in Fig. 1*b, d*. The main chain of part of strands I and F (open lines) and their partners related by the *r* dyad (solid

lines) are shown. Also shown are the side chains of Thr 245. *a*, T-state structure, direct interactions between the subunits. *b*, R-state structure, interactions mediated by a layer of water molecules.

and it shows well-preserved 222 symmetry (root-mean-square deviation for C_{α} atoms from an average structure = 0.30 Å). A few differences between subunits will be noted later, but generally the following discussion will refer to the structure of one subunit (labelled D): this subunit shows the clearest density; the critical parts of the molecule are not involved in crystal contacts.

Comparison of the R- and T-state structures

For the purposes of comparison, all motions will be considered as going from the R state to the T state. If the two structures are superimposed on their 222 axes, the major difference is seen to be a rotation of the AB dimer relative to the CD dimer by about 7° (Fig. 1*a*). The rotation axis is close to the molecular *p* dyad (the symmetry axis of the dimers), and there is no significant translation component along it (Table 1). Superposition of individual subunits rather than dimers improves the fit only marginally (Table 1), showing that the quaternary structure change is indeed a rotation of rigid dimers. This is a consequence of the much larger contact area between subunits within the dimer (2,190 Å² per subunit) compared to that between dimers (1,440 Å² per subunit).

The structural changes within the subunit were analysed systematically by superimposing in turn each secondary structure element, and calculating the deviations for each other element¹⁹. This analysis shows that only helices 8 and 9 and loops 6-F and 8-H shift significantly relative to the rest of the subunit (Fig. 1*d*). The large domain remains virtually unchanged, and it retains its position relative to the rigid core of the small domain. The rearrangement of loops 6-F and 8-H is discussed below. Helices 8 and 9 form the outer layer of the $\alpha\beta\alpha$ sandwich in the small domain: they perform a shear motion (net translation 1.2 Å) on the central β -sheet, but this movement does not affect the subunit contacts, nor do these helices carry any residues that bind ligands. The shear motion is accompanied by changes in the side-chain conformation of Glu 187 and Leu 205 and by slight variations in the side-chain torsion angles of the other residues at the helix-pair/ β -sheet interface, but the same side

chains remain in contact after the transition. The inner layer of the sandwich (helices 6, 7, 13) remains unchanged, because it is held in place by an extensive contact with the rigid large domain of the neighbouring (*p* related) subunit within the same dimer. The β -sheet is slightly tilted around an axis perpendicular to the sheet.

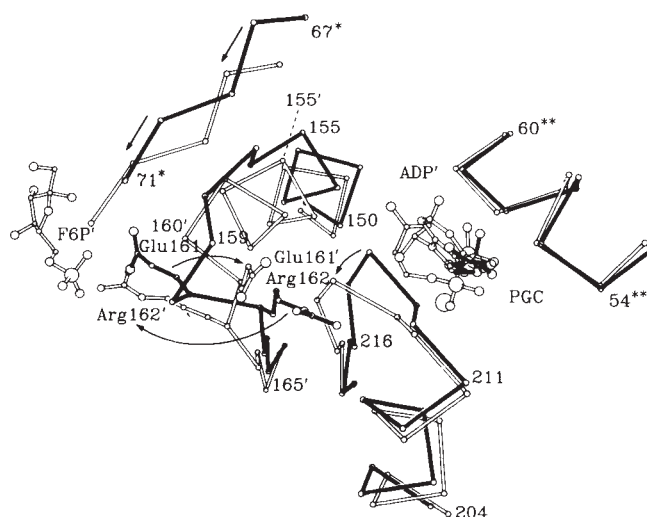


FIG. 3 Superposition of those parts of the T- (solid lines) and R-state (open lines) model that undergo large structural changes on the allosteric transition (8-H loop: around residues 211–216, 6-F loop: around 155–162) and its immediate surroundings. For comparison with Fig. 1*b*, the central (unstarred) subunit is D, residues of the *p*-dyad-related subunit C (belonging to the same dimer) are labelled with **, those of the *r*-dyad-related subunit A (other dimer) with *. Residues of the R-structure are distinguished by a prime. The T- to R-state transition is indicated by arrows. Also shown are the bound ligands: inhibitor 2-phosphoglycolate (PGC) for the T structure and the cooperative substrate F6P and the activator ADP for the R structure. Note the replacement in the Fru6P site of the R-state Arg162 by the T-state Glu 161, spoiling the binding of the phosphate group.

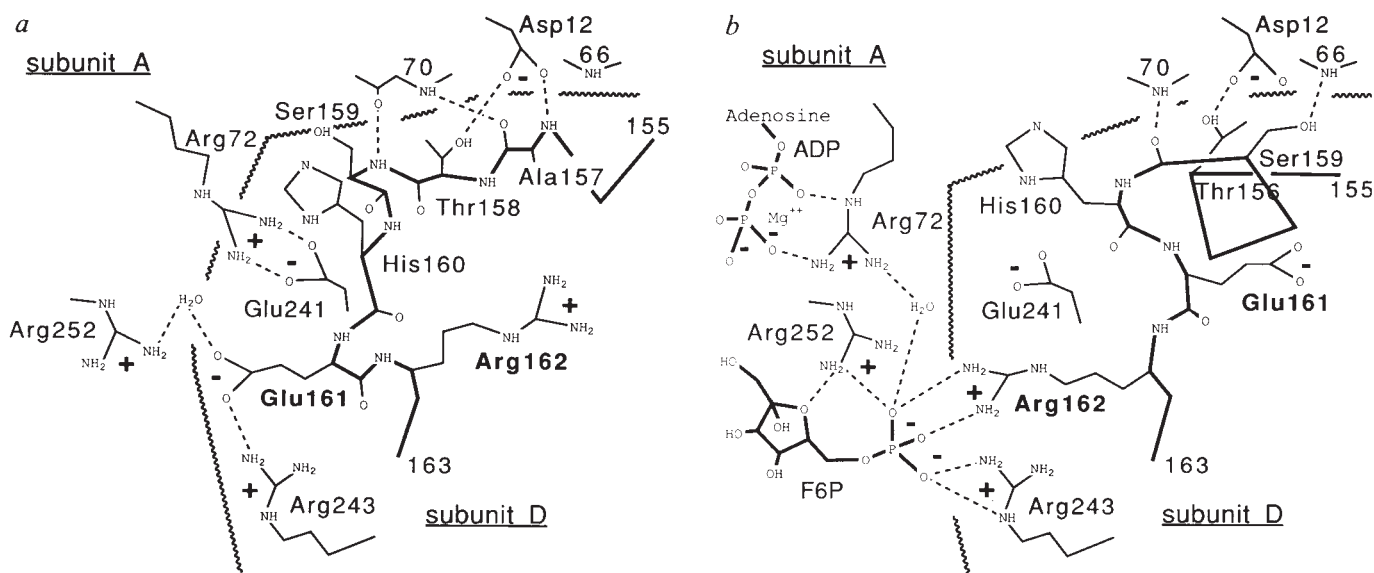


FIG. 4 Schematic view of the 6-F loop (155-162) showing its different interactions across the interface between dimers. The wavy line represents part of the boundary between *r*-dyad-related subunits (that is, the boundary

between dimers). *a*, T structure, showing Glu 161 in the Fru6P site, the Arg 72-Glu 241 salt bridge. *b*, R structure (with bound ligands ADP and Fru6P), showing the phosphate of Fru6P bound by Arg 162 and Arg 243.

Changes in subunit interfaces

We have detected no significant change at the large interface across the *p* axis within the dimer, even though the site which binds different types of effectors in the R and T structures lies there. One additional hydrogen bond between the subunits is formed in the T structure: Asp 59, which interacts with 03' of the ribose of ADP in the R structure (Fig. 5*c*), contributes to a hydrogen bond to His 215 in the neighbouring subunit (Fig. 5*b*).

On the other hand, the interface between the two dimers has to trigger and to accommodate their relative rotation in the allosteric transition. The enzyme has been designed to minimize structural changes at that interface by keeping the contact area small and by preserving interactions close to the quaternary rotation axis. There are two regions which change, in a passive and an active alteration at the interface. The passive change takes place around the *r* dyad, where the quaternary rotation from R to T closes the gap between symmetry-related β -sheet strands I. In the R structure, an ordered layer of water molecules bridges these strands (Fig. 2*b*, see also Fig. 12*b* in ref. 10 for the *E. coli* enzyme): in the T structure, these are expelled and direct hydrogen bonds (Fig. 2*a*) form across the interface. Two bonds around the *r* dyad are of the antiparallel β -sheet type: further away from the dyad, two hydrogen bonds between side chain (Thr 245) and main chain are formed. Other interactions

around the *r* dyad are close to the axis of the quaternary rotation and show little change.

The active change in the contacts between the dimers takes place at the end of helix 6 and the 6-F loop which packs against the loop 63-70 (C-4a). One side of the contact, the loop 6-F, changes its structure to compensate for the quaternary rotation, maintaining suitable contacts across the dimer interface. If the quaternary rotation to R took place without changing the structure of that loop the dimers would clash in this region of the interface; conversely, a T structure with the loop in the R conformation would leave a gap. In the T state the last turn of helix 6 (156-160) and the extended 6-F loop adopt a new conformation (Fig. 1*a,b*, and Fig. 3), that causes a critical change in the residues pointing into the adjacent Fru6P-binding site (see next section). The backbone of the C-4a loop that belongs to the large domain of the adjacent subunit remains unchanged. The R and T structures are held together by different sets of hydrogen bonds between polar groups of neighbouring dimers (Fig. 4). In the T state, part of the 6-F loop is stabilized by interactions of a parallel β -sheet type with another subunit (O₁₅₇-N₇₀, N₁₅₉-O₇₀; in the R state O₁₅₉ forms a bond with N₇₀ because the backbone has shifted by two residues). In the T state a new salt bridge, Arg 72-Glu 241 between the dimers is observed which may contribute to its higher stability (Figs 4,

TABLE 1 Overall fit of the T and R structure of PFK

Superimposed elements	Number of compared C α atoms	r.m.s.d.* ($\text{\AA}$)	r_p	R $\rightarrow$ T transformation†		θ	Screw component ($\text{\AA}$)
222 axes	1,276	1.70					
Dimer AB	638	0.93	0.99	0.02	0.14	3.1°	-0.1
Dimer CD	638	0.96	0.99	-0.04	0.13	-3.6°	0.1
Subunit A	319	0.85					
Subunit B	319	0.89					
Subunit C	319	0.86					
Subunit D	319	0.95					
Subunit D‡	237	0.43					

Overall fit between the C α atoms of the R and T-state structures, after superposition on the 222 axes or the specified subunits. Residues 152-162, 170, 171, 193-201, 213-216, 228-239 and 243-244 show deviations greater than the r.m.s. of the fit. The tetramer of the R structure has crystallographic 222 symmetry, whereas the 222 symmetry of the T structure is local. Note that the fit of the individual subunits is not significantly better than that of the dimers.

* Root-mean-square deviation.

† Transformation of dimer after superimposing the 222 axes, decomposed into a rotation by θ around an axis with the direction cosines r_p , r_q , r_r and a translation (screw component) along this axis. The slight departure of the rotation axis from the molecular *p* axis reflects the small deviations from exact 222 symmetry in the T-state crystals.

‡ Only C α of large domain (residues 1-137 and 252-304) and rigid core of small domain (helices 6 (143-151), 7 (177-185) and 13 (311-320) and β -sheet FGHI (164-168, 187-191, 216-221, 243-246)).

6). We found minor differences in the conformation and order of the 6-F loop of the different subunits in the crystal, but all these conformations show the same important differences from the R structure. The minor variations may merely manifest the loop's flexibility that is needed for the allosteric transition. The Arg 72-Glu 241 salt bridge is broken in two of the subunits, although the side chains still point towards each other: this is probably due to imperfect 222 symmetry in the crystals.

The active site

The ATP-binding site is formed almost entirely by residues of the large domain, whereas the Fru6P-binding site consists of residues from both domains and from the other dimer (Fig. 1c). Except for Arg 72, the structure of the ATP site remains

unchanged on the allosteric transition. In the R state, Arg 72 bridges the two substrates Fru6P and ATP, although it is never highly ordered: in the T state, Arg 72 forms a salt bridge with Glu 241, and cannot interact with the substrates (see Figs 4, 6).

On the other hand, the binding site for the cooperative substrate Fru6P is altered by the drastic change of the 6-F loop (see above) of the adjacent dimer. In the R state, a major part of the binding energy for Fru6P comes from Arg 162 and Arg 243 across the subunit interface, as shown by the reduced affinity (by factors of 165 and 53) for Fru6P in mutants lacking these side chains²⁰. On reorganization of this loop (Fig. 3, see also Fig. 4), Arg 162 swings around, bringing its guanidinium group close to the main chain of the end of helix 8. Glu 161, in turn, adopts the former (R state) position of Arg 162 and forms a

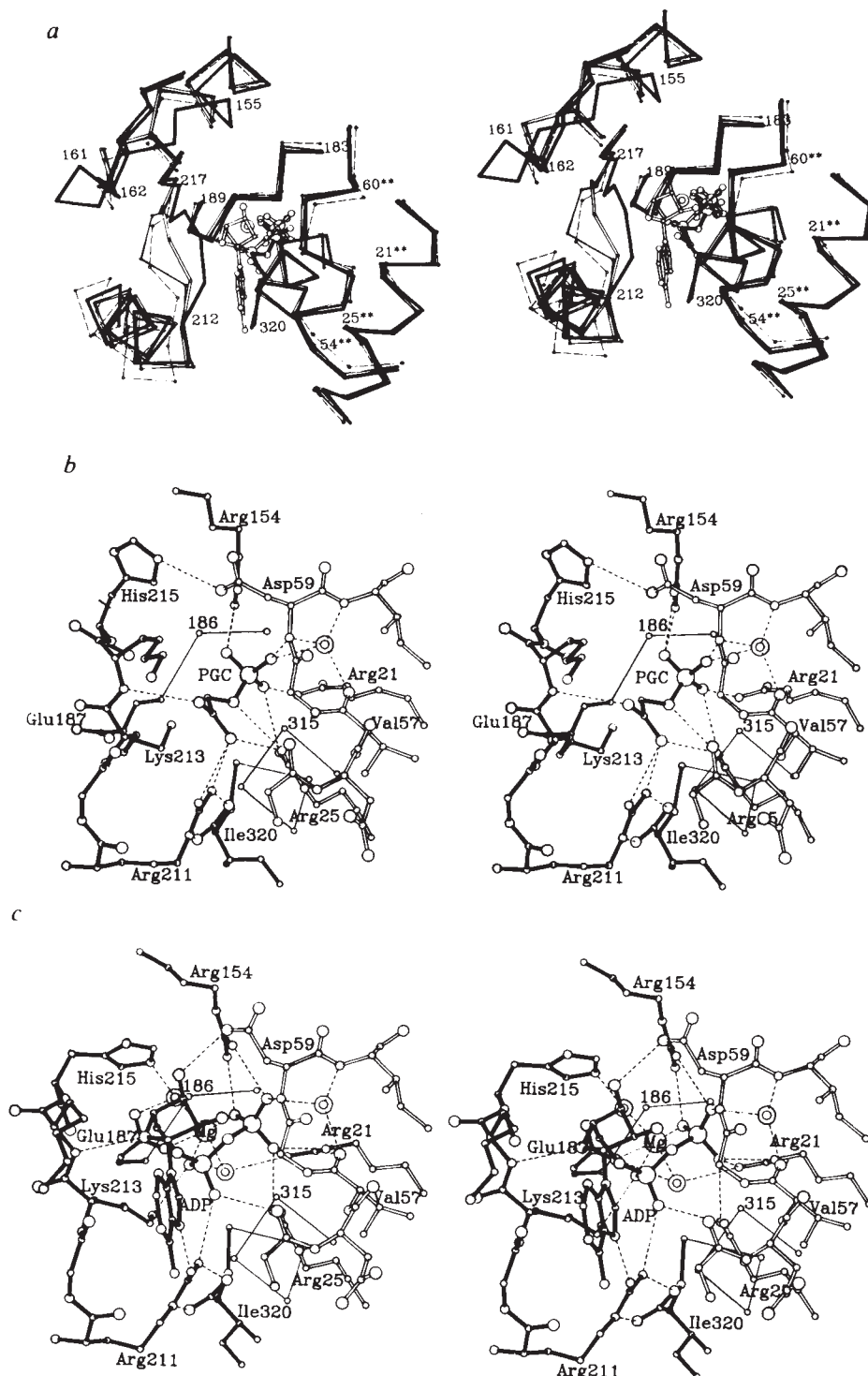


FIG. 5 Stereo views of the allosteric effector site. *a*, Superposition of C α -tracings around the effector site of several PFK structures, together with their bound effectors. Residues labeled with ** belong to the subunit C of the same dimer related by *p* dyad (subunit D unstarred). Residue 320 is the C terminus. Thick solid line, T state (*B. stearothermophilus* PFK with inhibitor 2-phosphoglycolate (PGC)); thick open line, R state (*B. stearothermophilus* PFK with activator ADP); thin line, R state (*B. stearothermophilus* PFK with inorganic phosphate bound to the effector site^{8,9}); broken line, R state (*E. coli* PFK, unliganded¹¹). Only the 8-H loop (213–216) shows major movement in response to the change of ligand. *b*, Effector site of the T structure with bound inhibitor PGC. The two subunits related by the *p* dyad are distinguished by open (C) and solid lines (D). Thin lines: C α tracings. The carboxyl group of PGC pinches together the lips of the site. *c*, Effector site of the R structure with bound activator ADP. Same representation as in *b*. The larger ADP causes a widening of the site compared to PGC, but the binding of the ADP β -phosphate and the PGC phosphate are similar.

salt bridge with Arg 243; its carboxylate group overlaps with the old guanidinium position; the different courses of the backbones compensate partly for the different lengths of the glutamate and arginine residues. In this way the positive charges of arginines 162 and 243 are no longer available for binding to the doubly negative charged 6-phosphate of Fru6P. This accounts for the large decrease in the affinity for Fru6P (in *E. coli* PFK, residue 161 is Gln, so the charge change is smaller).

The analysis of the T structure at low resolution¹² indicated that closing of the Fru6P-binding site inhibits the enzyme. Quaternary and tertiary changes (in particular, movement of loop 9-I and strand I) do indeed combine to narrow this site. Superposition of the active sites of the two states reveals that a Fru6P molecule positioned as in the R structure would clash with the protein in the T state: arginines 243 and 252 and Glu 222 approach the substrate to within 1.9 Å. These steric barriers might be overcome by minor readjustments of some side chains, but their effect on Fru6P binding is difficult to assess. Our comparison of the active sites suggests that reorganization of arginines 72 and 162, rather than closure of the binding site, underlies the low Fru6P affinity of the T structure.

The effector site

Figure 5a shows the C_{α} tracings of the effector sites of all *B. stearothermophilus* PFK structures⁹ and of the unliganded *E. coli* PFK structure¹¹ superimposed. Apart from the 8-H loop (residues 213–215), most of the site remains unchanged in response to binding of effectors of different kinds. The movements of the 8-H loop may be described as a hinge motion with its pivots at residues 212 and 216. With no ligands, or with phosphate, the site is open. When the effectors ADP of the R structure or 2-phosphoglycolate of the T structure are bound, the cleft closes up to pack tightly around them. The position of the 8-H loop is then controlled by the different sizes of the effectors, the smaller, 2-phosphoglycolate, inducing the narrowest cleft. The change in the 8-H loop structure appears to be correlated with the reorganization of the adjacent 6-F loop: in particular, the amino termini of the two β -strands F and H are shifted in concert, tied together by hydrogen bonds $O_{162}-N_{216}$ and $N_{163}-O_{216}$.

Figure 5b, c juxtaposes the details of the effector site in the T and R structures. The modes of binding of the inhibitor and activator are strikingly similar, with the terminal phosphates bound in the same way to arginines 154, 21 and 25, deep at the bottom of the effector-binding cleft. Arginines 25 and 211 which fix the α -phosphate of ADP in the R structure interact with the bridging oxygen and the carboxylate group of 2-phosphoglycolate in the T structure. The main-chain amide of residue 214 which forms a hydrogen bond with O_4' of the ADP ribose, interacts with the 2-phosphoglycolate carboxyl in the T structure. The 2-phosphoglycolate molecule is almost totally buried, its carboxylate group just reaching the surface and stapling the two rims of the cleft together. Model building shows that the extra

methylene group of PEP can easily be accommodated in the site. The only major difference in the effector site is the conformation of the side chain of Glu 187 (Fig. 5b, c), which is coordinated to the magnesium ion of ADP in the R structure, and is rotated away from the ligand in the T structure. Its χ_1 torsion angle changes by $\sim 140^\circ$, and the side chain adopts a new position sandwiched between the central β -sheet of the small domain and helix 9. To make room, the side chain of Leu 205 also changes its conformation (data not shown).

The allosteric mechanism

On transition from the R to the T state, the rigid dimers perform a contrary rotation around their common dyad axis. This motion couples the changes in the four binding sites for the cooperative substrate Fru6P, which lie between dimers. The rotation of the dimers is accommodated by removal of the layer of water molecules that lies between them in the R structure, and by the restructuring of the 6-F loop. Intermediate states in these contacts would be unstable: all the waters must be either present or absent; and the alternative contacts made by the 6-F loop with the neighbouring subunit would be disrupted in an intermediate conformation. The nature of these changes suggests that the transition is concerted, with the 222 symmetry being maintained.

The removal or insertion of the water layer between the dimers may represent a general mechanism of alternative packing between or within protein molecules. Any complementary array of hydrogen-bonding groups may be pulled apart, and a layer of waters inserted between them, forming an energetically favourable structure, though with some loss of entropy. But in the allosteric transition described here, this change may be passive, allowing the transition but not triggering it.

On the other hand, the rearrangement of the 6-F loop couples the binding of Fru6P to the quaternary structure change, and to the binding of heterotropic effectors. Replacement of Arg 162 in the Fru6P site by Glu 161 is sufficient to explain the low affinity of the T state for Fru6P, but narrowing of the binding site may also play some part. Surprisingly, it seems unimportant that the effector site lies between subunits in the dimer. Despite the binding of effectors with opposite allosteric effects, most of the site remains unchanged. Only the 8-H loop moves in response to a change of effector, but this movement is coupled to the rearrangement of the 6-F loop. It is not obvious how the motions are coupled, but the close linkage of the ends of the two loops to the sheet strands F and H may be important. Presumably, the low-affinity T-state conformation can only be formed when the lips of the effector site are closed: this closure is induced by inhibitors, but cannot happen when the larger activator ADP is bound, hence their opposite effects. PFK is very economical in using the same conformational change to trigger the allosteric transition in both homotropic and heterotropic interactions, with minimal tertiary structure changes coupled to the quaternary structure change. □

Received 18 August; accepted 16 November 1989.

- Monod, J., Wyman, J. & Changeaux, J. P. *J. molec. Biol.* **12**, 88–118 (1965).
- Koshland, D. E., Nemethy, G. & Filmer, D. *Biochemistry* **5**, 365–385 (1966).
- Perutz, M. F., Fermi, G., Luisi, B., Shaanan, B. & Liddington, R. C. *Accounts chemical Res.* **20**, 309–321 (1987).
- Kantrowicz, E. R. & Lipscomb, W. N. *Science* **241**, 669–674 (1988).
- Barford, D. & Johnson, L. N. *Nature* **340**, 609–616 (1989).
- Blangy, D., Buc, H. & Monod, J. *J. molec. Biol.* **31**, 13–35 (1968).
- Valdez, B. C., French, B. A., Younathan, E. S. & Change, S. H. *J. biol. Chem.* **264**, 131–135 (1989).
- Evans, P. R. & Hudson, P. J. *Nature* **279**, 500–504 (1979).
- Evans, P. R., Farrants, G. W. & Hudson, P. J. *Phil. Trans. R. Soc.* **B293**, 53–62 (1981).
- Shirakihara, Y. & Evans, P. R. *J. molec. Biol.* **204**, 973–994 (1988).
- Rypniewski, W. R. & Evans, P. R. *J. molec. Biol.* **207**, 805–821 (1989).
- Evans, P. R., Farrants, G. W. & Lawrence, M. C. *J. molec. Biol.* **191**, 713–720 (1986).

- French, B. A. & Chang, S. H. *Gene* **54**, 65–71 (1987).
- Hellings, H. W. & Evans, P. R. *Eur. J. Biochem.* **149**, 363–373 (1985).
- Jack, A. & Levitt, M. *Acta Crystallogr.* **A34**, 931–935 (1978).
- Hendrickson, W. A. & Konnert, J. in *Biomolecular Structure, Function, Conformation and Evolution* Vol. 1 (ed. Srinivasan, R.) 43–47 (Pergamon, Oxford, 1981).
- Bricogne, G. *Acta Crystallogr.* **A32**, 832–847 (1976).
- Read, R. J. *Acta Crystallogr.* **A42**, 140–149 (1986).
- Lesk, A. M. & Chothia, C. *J. molec. Biol.* **174**, 175–191 (1984).
- Berger, S. A. & Evans, P. R. *Nature* (in the press).

ACKNOWLEDGEMENTS. T.S. was supported by an EMBO fellowship. We thank Max Perutz, Stuart Berger and Craig Kundrot for comments. The coordinates and diffraction data will be deposited in the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, New York 11973, USA.

A possible new association of a pulsar with a supernova remnant

Namir E. Kassim & Kurt W. Weiler

E. O. Hulburt Center for Space Research, Naval Research Laboratory,
Code 4035, Washington, DC 20375-5000, USA

THE link between supernova explosions and the formation of neutron stars seems to be rather well established, even if only type II supernovae are expected to leave a stellar remnant¹. However, only five examples of pulsar/supernova remnant associations are known in our Galaxy, and one in the Large Magellanic Cloud. Notwithstanding a further thirty examples^{1,2} where plerionic radio or non-thermal X-ray emission indirectly implies the presence of an active neutron star, each possible pulsar/supernova remnant association remains interesting and useful. Here we report new measurements at 90-cm wavelength, which show that the W30 supernova remnant G8.7-0.1 is large, roughly shell-shaped, and more extended than previously thought. PSR1800-21, one of the youngest galactic pulsars, seems to lie within the western extension of the remnant. At a distance determined to be ~ 5.5 kpc, G8.7-0.1 seems to be close to PSR1800-21 in space as well as on the sky, and estimates³⁻⁵ indicate that the two objects are of similar age. Thus, a physical relation seems possible, but raises the dynamical problem of having a simultaneously born pulsar travel to a position near the edge of a supernova remnant within its visible lifetime. We discuss possible solutions to this difficulty.

Odegard⁵ discusses a possible association of the W30 supernova remnant (SNR) (G8.7-0.1) with the nearby pulsar PSR1800-21. His map of G8.7-0.1 at 57.5 MHz is strongly affected by absorption from foreground H II regions, but shows a finger of emission that extends towards the (then) best available pulsar position. He obtains an estimated age for G8.7-0.1 of $\sim 15,000$ yr from surface brightness-age ($\Sigma-t$) relations³ and a best distance estimate of ~ 2.5 kpc from several surface brightness-linear diameter ($\Sigma-D$) relations^{3,6,7}. Clifton and Lyne⁴ place PSR1800-21 at a distance of ~ 5.3 kpc, using its dispersion measure, and assign it a characteristic age of $\sim 16,000$ yr on the basis of its period (P) and period derivative ($\dot{P}$). Thus, although G8.7-0.1 and PSR1800-21 are close on the sky and of similar ages, the difference in the distance estimates meant that a spatial association was unlikely. The problem was compounded by the large space velocity necessary for PSR1800-21 to reach its present position, which is quite distant from the centre of the 57.5-MHz emission from G8.7-0.1, within $\sim 15,000$ yr. Thus, Odegard concluded that an association seemed improbable but could not be ruled out.

Very-large-array (VLA) observations⁸ of PSR1800-21 at 20 cm show that the position given by Clifton and Lyne and used by Odegard is in error. But, although the pulsar was definitely detected and its position determined with high accuracy, no possible SNR association was identified and the nature of the faint extended emission seen on the VLA 20-cm map was not determined.

We have mapped the W30 complex at a wavelength of 90 cm (330 MHz) using the National Radio Astronomy Observatory VLA in the D configuration (maximum baseline ~ 1 km; resolution $\sim 3'$). Although the region is a complicated mixture of thermal and non-thermal emitters requiring a detailed analysis presented elsewhere⁹, our relatively low observing frequency and high brightness sensitivity allowed us to obtain the most accurate map so far of the non-thermal emission from G8.7-0.1. Shorter-centimetre-wavelength single-dish measurements¹⁰⁻¹² are too heavily dominated by thermal emission and galactic background structure, shorter-centimetre-wavelength interferometer measurements⁸ have too high a resolution and too low a brightness sensitivity, and longer-metre-wavelength interferometer measurements⁵ are too heavily affected by thermal absorption.

The region around the W30 complex is shown in Fig. 1; the position of PSR1800-21 is indicated by a cross. G8.7-0.1 is a large, roughly shell-shaped SNR which is more extended than previously⁵ realized and PSR1800-21 lies within its outer extensions. At this frequency, the thermal component contributes $<15\%$ of the observed flux density.

PSR1800-21 is a fast, young pulsar and is one of only six galactic pulsars with a characteristic age under 20,000 yr. As can be seen in Table 1, half of these youngest galactic pulsars have already been associated with known SNRs.

It has been shown⁹ that several of the H II regions in the W30 complex are visible in absorption at low frequencies against the non-thermal emission of G8.7-0.1 and are therefore closer than the SNR. As previous observations¹³⁻¹⁶ of recombination line emission reliably place these H II regions at $\sim 5-6$ kpc, a new estimate for the distance to G8.7-0.1 is $\geq \sim 5$ kpc. The lack of absorption along the line of sight to G8.7-0.1 by yet another W30 H II region at the same distance implies a weaker upper limit of $\leq \sim 6$ kpc. This places G8.7-0.1 at $\sim 5.5 \pm 1$ kpc, considerably farther than the previous estimate. Our estimate is independently supported by OH absorption spectra along the line of sight to the SNR¹⁷ which also show features compatible with the distance to G8.7-0.1 being $\sim 5-6$ kpc. At this distance, to within the uncertainties, G8.7-0.1 seems to be spatially, as well as on the sky, close to PSR1800-21 so that the likelihood of an association is enhanced. The best estimates of the properties of the two objects and their relation are summarized in Table 2.

The dynamical problem, however, remains. At a distance of ~ 5.5 kpc, if we accept an age of $\sim 15,000$ yr, the average expansion velocity of G8.7-0.1 is $\sim 2,700$ km s⁻¹. The required transverse velocity of PSR1800-21, if it was born in the apparent centre of symmetry of G8.7-0.1, is $\sim 1,700$ km s⁻¹. As supernovae are known to have initial expansion velocities of up to 20,000 or 30,000 km s⁻¹ (and even older, decelerated remnants such as the Cygnus Loop still show filamentary motions of a few hundred km s⁻¹) such an average expansion velocity for G8.7-0.1 is not unreasonable. Such a large transverse velocity for PSR1800-21, however, would be unprecedented.

Pulsars are long-lived objects with ages of up to a few million years and high average velocities ($\langle v^2 \rangle^{1/2} \sim 200$ km s⁻¹)¹⁸ so they

TABLE 1 The youngest known galactic pulsars

Pulsar	0531+21	0833-45	1509-58	1737-30*	1800-21*	1823-13*
d (kpc)	2	0.5	4	3	5	5
P (ms)	33	89	150	607	133	101
$\dot{P}$ (10^{-15} s s ⁻¹)	423	125	1540	466	134	72
$\dot{E}$ (10^{36} erg s ⁻¹)	460	7	18	0.07	1	2
Characteristic age (10^3 yr)	1.2	11	1.6	20	16	22
SNR	Crab Neb.	Vela XYZ	MSH15-52	None	W30SNR	None
Galactic name	G184.6-5.8	G263.9-3.3	G320.4-1.2		G8.7-0.1	

* Data from ref. 8.

TABLE 2 Properties of a possible G8.7-0.1/PSR1800-21 association

G8.7-0.1	
Centre	
Right ascension (1950)	18 ^h 02 ^m 20 ^s
Declination (1950)	-21° 34'
Distance (kpc)	~5.5 ± 1
Age (yr)	~15,000
Radius (min)	~26
Radius (pc)	~42
Average expansion velocity (km s ⁻¹)	~2,700
PSR1800-21	
Right ascension (1950)*	18 ^h 00 ^m 51.09 ^s
Declination (1950)*	-21° 37' 17.5"
Distance (kpc)	~5.3
Age (yr)	~16,000
Distance from G8.7-0.1 centre (min)	~19
Distance from G8.7-0.1 centre (pc)	~29
Average velocity from G8.7-0.1 centre (km s ⁻¹)	~1,700

* Data from ref. 8.

will eventually exceed the boundaries of any simultaneously created SNR. To do this within ~15,000 yr, however, seems unlikely because the highest observed transverse motion for a pulsar¹⁸ is <400 km s⁻¹. Also, if PSR1800-21 is a typical neutron star with a mass of 1.4 $M_{\odot}$ and moving at a velocity of 1,700 km s⁻¹, its minimum kinetic energy of 4×10^{49} erg is 4% of the canonical 10^{51} erg of kinetic energy released in a supernova explosion, a disturbingly large fraction.

Because we are not able to detect PSR1800-21 clearly in our 90-cm map, we can place an upper limit on its flux density of $S_{90\text{ cm}} < 250$ mJy. Combining this with measurements^{4,8} at higher frequencies implies a spectral index, $\alpha \geq -1.9$ ($S \propto \nu^{\alpha}$ where ν is the frequency), a fairly normal value for a pulsar¹⁹. Assuming $\alpha = -2$ we can extrapolate its flux density to the standard frequency of 400 MHz and find that only a few pulsars are known that are as luminous as PSR1800-21 ($L_{400\text{ MHz}} \sim 6 \times 10^3$ mJy kpc²). At a density²⁰ of $< 10^{-2}$ kpc⁻², the *a posteriori* probability of having PSR1800-21 within 50 pc of the centre of G8.7-0.1 is $< 10^{-4}$. Even accepting the limitations of

a posteriori calculations, the probability of having such a luminous pulsar in any given 1° segment of galactic longitude, at any latitude or distance, is only ~2%. Thus, we accept an association between PSR1800-21 and G8.7-0.1 as possible and consider possible causes for the remaining inconsistencies.

(1) The distance and/or age estimate for G8.7-0.1 could be incorrect. Although the age estimate for G8.7-0.1 is based on one of the weak surface brightness/physical parameter relations for SNRs, the distance is determined⁹ from the much more firmly established H II region distances¹³⁻¹⁶. Thus, an incorrect age for G8.7-0.1 is a possibility although most models would place the age of such a large, but still relatively luminous, SNR between 10,000 and 50,000 yr.

(2) The distance and/or age estimate for PSR1800-21 could be incorrect. Distances determined by dispersion measure are dependent on the average thermal-electron density $\langle n_e \rangle$ along the line of sight and the W30 complex contains a number of H II regions of obviously higher density than the average interstellar medium. As most of the H II regions in the direction of W30 are at ~5-6 kpc, however, it is difficult to place PSR1800-21 closer and still behind an H II region. Placing it farther away makes it more luminous and even less common in the Galaxy. Thus, it is likely to be at ~5.3 kpc as indicated by its dispersion measure. The true age for PSR1800-21 is also somewhat uncertain. Studies that show the characteristic age (τ_c) of pulsars to be generally of the order of, or older than, the true age (τ_i) deal almost entirely with much older pulsars having $\tau_c = 0.5 P \dot{P}^{-1} > 100,000$ yr. Although the other examples of young pulsars given in Table 1 imply that $\tau_c \approx \tau_i$, the age of PSR1800-21, as that for G8.7-0.1, remains a weak link in a possible association. Accepting an association but postulating true ages that are closer to, for example, ~100,000 yr removes most physical objections but requires that the age estimates of both the pulsar and the SNR be far too low and by the same amount. That, again, seems unlikely. Thus, we consider it at least possible that PSR1800-21 has a very high transverse velocity of ~1,700 km s⁻¹ or a proper motion of ~0.1" yr⁻¹. Such a hypothesis can be tested by normal radio-astronomy techniques in a few years.

(3) G8.7-0.1 could be very large. An alternative, and perhaps more promising possibility, is that G8.7-0.1 is much larger than our measurements can detect (sizes > 1° are strongly attenuated

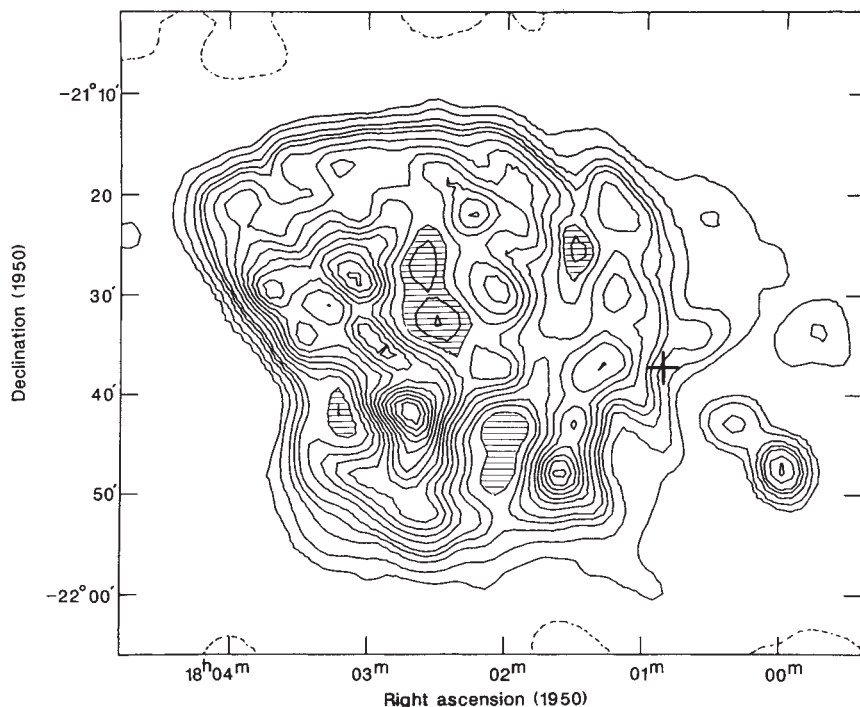


FIG. 1 VLA 90-cm map of the W30 complex with a $3.7' \times 3.0'$ (at position angle 32.9°) (full width at half maximum; FWHM) synthesized beam. The field has been corrected for the primary beam (~156' FWHM) attenuation. The large, roughly shell-shaped SNR G8.7-0.1 is evident and the position of the pulsar PSR1800-21 is indicated by a cross. G8.7-0.1 is considerably larger than previously realized, extending beyond the position of PSR1800-21. The contour levels are -0.125, 0.063, 0.125, 0.250, 0.375, 0.50, 0.75, 1.00, 1.25... 3.25, 3.50 Jy per beam. Hatched areas are local depressions. The negative contour is dashed. The bright small-diameter source near right ascension = 18^h 00^m, declination = -21° 48' is a known H II region.

even at this smallest configuration and lowest observing frequency of the VLA), as faint extensions in the direction of PSR1800–21 indicate. Such a solution has been proposed²¹ to explain the low proper motion²² of PSR0833–45 and its offset from the traditional centre of Vela XYZ. A larger and asymmetric shape for G8.7–0.1 could place the true centre of the SNR much closer to the pulsar position and lower its required transverse velocity. There is some evidence for such low-level emission in Fig. 1 and, even with confusion from the galactic background and a relative insensitivity to steep-spectrum, non-thermal emission, there is also evidence for a greater extension on the centimetre-wavelength single-dish maps of the W30 region^{10–12}.

Measurements of the pulsar proper motion and new, low-frequency, high-resolution, filled-array searches for extended non-thermal emission are needed to confirm or discount the possible association between G8.7–0.1 and PSR1800–21 indicated by our data. □

Received 17 April; accepted 13 November 1989.

1. Weiler, K. W. & Sramek, R. A. *Rev. Astr. Astrophys.* **26**, 295–341 (1988).
2. Seward, F. D. *Space Sci. Rev.* (in the press).
3. Caswell, J. L. & Lerche, I. *Mon. Not. R. astr. Soc.* **187**, 201–216 (1979).
4. Clifton, T. R. & Lyne, A. G. *Nature* **320**, 43–45 (1986).
5. Odegard, N. *Astr. J.* **92**, 1372–1380 (1986).
6. Milne, D. K. *Aust. J. Phys.* **32**, 83–89 (1979).
7. Huang, Y.-L. & Thaddeus, P. *Astrophys. J.* **295**, L13–L16 (1985).
8. Braun, R., Goss, W. M. & Lyne, A. G. *Astrophys. J.* **340**, 355–361 (1989).
9. Kassim, N. E. & Weiler, K. W. *Astrophys. J.* (in the press).
10. Reich, W., Fürst, E., Steffen, P., Reif, K. & Haslam, C. G. T. *Astr. Astrophys., Suppl.* **58**, 197–248 (1984).
11. Altenhoff, W. J., Downes, D., Pauls, T. & Schraml, J. *Astr. Astrophys., Suppl.* **35**, 23–54 (1978).
12. Handa, T., Sofue, Y., Naomasa, N., Hirabayashi, H. & Inoue, M. *Publ. astr. Soc. Japan* **39**, 709–753 (1988).
13. Downes, D., Wilson, T. L., Biegling, J. & Wink, J. *Astr. Astrophys. Suppl.* **40**, 379–394 (1980).
14. Lockman, F. J. *Astrophys. J., Suppl.* (in the press).
15. Wilson, T. L. *Astr. Astrophys.* **19**, 354–368 (1972).
16. Georgelin, Y. M. & Georgelin, Y. P. *Astr. Astrophys.* **49**, 57–79 (1976).
17. Turner, B. E. *Astr. Astrophys., Suppl.* **37**, 1–332 (1979).
18. Lyne, A. G., Anderson, B. & Salter, C. J. *Mon. Not. R. astr. Soc.* **201**, 503–520 (1982).
19. Manchester, R. N. & Taylor, J. H. *Pulsars* (Freeman, San Francisco, 1977).
20. Lyne, A. G., Manchester, R. N. & Taylor, J. H. *Mon. Not. R. astr. Soc.* **213**, 613–639 (1985).
21. Seward, F. D. in *Imaging X-ray Astronomy* (ed. Elvis, M.) (Cambridge Univ. Press, in the press).
22. Bignami, G. F. & Caraveo, P. A. *Astrophys. J.* **325**, L5–L7 (1988).

(luminosity in the range 10^{34} – 10^{38} erg s^{−1}) at a distance of the order of the galactic scale (>10 kpc) can be detected with a single scan.

So far, we have discovered about a dozen hard-X-ray sources from the galactic inner disk⁵. We have conducted several scans over these new sources separated typically by more than six months. Most of the new sources were found to be below the detection limit in the repeat scan, so the new X-ray sources are probably transient or highly variable. Six of the new sources are located in the narrow region near $l \approx 30^\circ$ and the others are scattered more uniformly over the entire galactic inner disk⁵. The positions of these six new sources are given in Fig. 1 together with the catalogued X-ray source Sct X-1. Two of the transient sources (No. 5 and No. 7) have been cross-scanned to determine their positions accurately. Pointing observations of these two sources and Sct X-1 revealed clear pulsations of 94.8 s for No. 5 (refs 1, 6), 29.5 s for No. 7 (refs 1, 7) and 111 s for Sct X-1 (refs 1, 8). We also discovered an 81.1-s pulsation at the 4.5σ level from the $1^\circ \times 2^\circ$ (full width at half maximum) region that includes source No. 3. Because of a crowded observation schedule no pointing observations of the other three sources were carried out and we are still not certain whether the other three sources are also X-ray pulsars. We suggest, however, that they are, because the X-ray spectra can be fitted with a power-law model using a photon index of 0.4–2.2 and the iron emission line, the same sort of model used to fit the spectra of catalogued X-ray pulsars. In Table 1 we summarize the observations and spectral fit. If all seven sources are X-ray pulsars, the concentration of X-ray pulsars in the region of $l \approx 30^\circ$ is higher than in any other regions of the galactic plane (see ref. 1).

The region around $l \approx 30^\circ$ has a conspicuous intensity shoulder in infrared, radio, CO molecular line, γ -ray and other measurements⁹, and it is often referred to as the 5-kpc arm². Because the newly discovered X-ray pulsars (and candidates) are concentrated in the $l \approx 30^\circ$ region, these new X-ray sources are probably associated with the 5-kpc arm. To confirm this, we need distance information. As no optical identification of these new X-ray sources is available, we estimate the distance using our X-ray data. The observed hydrogen number density N_H for most of the new X-ray sources is $\sim 10^{23}$ cm^{−2} (Table 1). Furthermore, we found that N_H values were nearly constant during the

Is the 5-kpc galactic arm a colony of X-ray pulsars?

K. Koyama, M. Kawada, H. Kunieda, Y. Tawara, Y. Takeuchi & S. Yamauchi

Department of Astrophysics, Nagoya University, Furo-cho Chikusa-ku, Nagoya, 464, Japan

MORE than two dozen binary X-ray pulsars have been discovered in our Galaxy¹. Because of instrumental constraints, however, most of these have been discovered within only a few kiloparsecs (kpc) of the Sun. Here we report the discovery by the Ginga satellite of four X-ray pulsars and three hard X-ray sources in the galactic plane at galactic latitude $l \approx 30^\circ$ and a distance of about 10 kpc (the '5-kpc arm'²). This indicates that the 5-kpc arm has been the site of vigorous formation of X-ray pulsars, and that in general their distribution might be expected to be non-uniform.

We have been conducting a systematic search for transient pulsars by scanning the galactic plane using the large-area counters (LAC)³ onboard Ginga⁴. These observations were made on several occasions between April 1987 and April 1989, covering the galactic inner disk ($|l| < 60^\circ$) uniformly. The detection limit for the discovery of a new X-ray source in a single galactic-plane scan is better than 10^{-11} erg s^{−1} cm^{−2} in the 2–37-keV band. Therefore, most typical X-ray pulsars

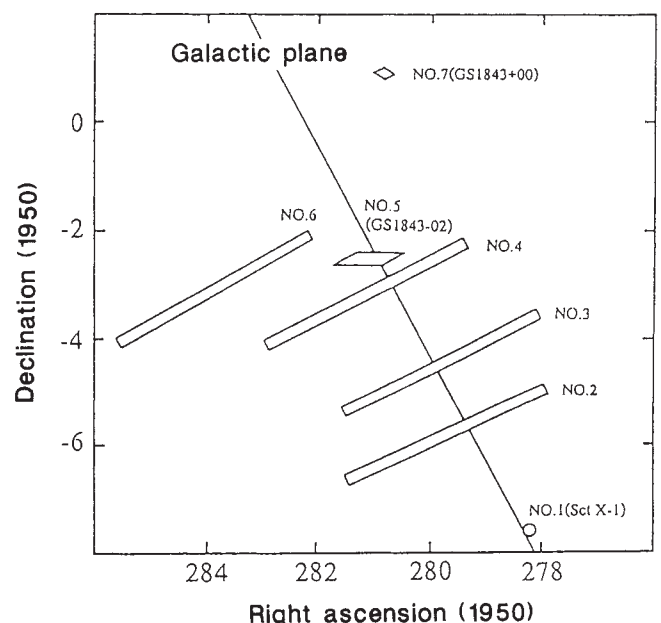


FIG. 1 The 90% error regions of the new transient pulsars and hard transient sources.

TABLE 1 Summary of the newly discovered X-ray pulsars and hard X-ray transients

No.	Name	Position		Pulse period (s)	Intensity in range 2–10 keV (10^{-11} erg s $^{-1}$ cm $^{-2}$)	Power-law index	log N_H (cm $^{-2}$)
		1	b				
1	Sct X-1	24.5 ± 0.01	-0.2 ± 0.01	111	4	1.8 ± 0.1	23.4 ± 0.4
2	—	26.6 ± 0.2	$1.5 \sim -2.5$	—	2	1.5 ± 0.3	22.7 ± 0.3
3*	—	27.9 ± 0.1	$-1.9 \sim 2.1$	81.1†	6	1.9 ± 0.2	22.8 ± 0.2
4*	—	29.7 ± 0.1	$-2.5 \sim 1.5$	—	2	1.8 ± 0.1	22.9 ± 0.1
5	GS1843–02	30.2 ± 0.2	1.1 ± 0.2	94.8	14	1.2 ± 0.1	23.4 ± 0.2
6*	—	31.3 ± 0.1	$-4.7 \sim -0.7$	—	4	2.2 ± 0.2	23.7 ± 0.1
7	GS1843+00	33.1 ± 0.1	1.7 ± 0.1	29.5	80	0.4 ± 0.1 ‡	< 22.0

* From ref. 10.

† Detection at the 4.5σ level.

‡ Power law plus exponential cutoff.

pointing observations. This is further evidence for the hypothesis that the observed N_H is mainly interstellar, rather than circumstellar, matter. Hayakawa *et al.*⁹ estimated that the N_H value to $l \approx 30^\circ$ at 10 kpc distance is $\sim 10^{23}$ cm $^{-2}$. Therefore the distances of new X-ray sources are likely to be ~ 10 kpc (± 4 kpc). Thus, we suggest that the new X-ray sources are not only distributed toward the $l \approx 30^\circ$ region but also located in the 5-kpc arm. At 10 kpc, the X-ray luminosities from the new X-ray sources are estimated to be 10^{35} – 10^{37} erg s $^{-1}$, well within the range of observed luminosities from X-ray pulsars.

In Fig. 2 we plot the spatial distribution in the galactic plane of the catalogued X-ray pulsars together with the region in which the newly discovered X-ray sources are mostly likely to lie. This shows the possible concentration of these sources in the 5-kpc arm.

The strong variability of these new X-ray sources indicates that they are transient X-ray pulsars in a Be-star binary system¹⁰. The typical age of a Be-star binary pulsar is estimated to be of the order of 10^7 yr (ref. 11). Also, the small scale height derived from near-infrared observations indicates that the 5-kpc arm consists of young stars with ages of the order of 10^7 yr (ref. 12). Therefore, these newly discovered X-ray pulsars (and candidates) in the 5-kpc arm may originate from the same star-forming episode revealed by the infrared observations. Thus, a systematic survey for new X-ray pulsars provides a method for investigating

the structure of the galactic arm as well as for studying the evolution of X-ray pulsars. □

Received 4 July; accepted 21 November 1989.

1. Nagase, F. *Publ. Astr. Soc. Japan* **41**, 1–79 (1989).
2. Burton, W. B. *A. Rev. Astr. Astrophys.* **14**, 275–306 (1976).
3. Turner, M. J. L. *et al. Publ. Astr. Soc. Japan* **41**, 345–372 (1989).
4. Makino, F. *Astr. Lett. Commun.* **25**, 223–233 (1987).
5. Koyama, K. in *Physics of Neutron stars and Black Holes* (ed. Tanaka, Y.) 55–65 (Universal Academy Press, Tokyo, 1988).
6. Makino, F. & the Ginga team, *IAU Circ. No.* 4661 (1988).
7. Makino, F. & the Ginga team, *IAU Circ. No.* 4587 (1988).
8. Makino, F. & the Ginga team, *IAU Circ. No.* 4679 (1988).
9. Hayakawa, S. *et al. Astr. Astrophys.* **58**, 325–330 (1977).
10. Koyama, K. *et al. Publ. Astr. Soc. Japan* **41**, 483–490 (1989).
11. Rappaport, S. & van den Heuvel, E. P. J. in *Be Stars* (eds Jaschek, M. & Groth, H.) 327–346 (Reidel, Dordrecht, 1982).
12. Hayakawa, S. *et al. Astr. Astrophys.* **100**, 116–123 (1980).

ACKNOWLEDGEMENTS. We thank C. Day and J. Deeter for comments on the manuscript.

Peak-current variations of lightning return strokes as a function of latitude

Richard E. Orville

Department of Atmospheric Science, State University of New York, Albany, New York 12222, USA

LATITUDINAL variations in lightning-flash characteristics have been hypothesized^{1,2}, but have not been found to exist^{3,4}. This may be because lightning characteristics are constant over the Earth, or because the observations have been made over too small an area of the globe. To investigate the question of latitudinal variation, I have used a network of thirty-six gated wideband magnetic direction finders covering the eastern United States to measure the characteristics of cloud-to-ground lightning flashes. These characteristics include the flash location, the flash polarity, the number of strokes in the flash and the peak magnetic radiation field in the first return stroke of the flash. From the peak radiation field, I estimate the peak current⁵. More than five million lightning ground flashes were recorded by the network in 1988. An examination of the mean peak current in the first return stroke shows that the peak current varies by almost a factor of two, from 25 kA in New England to 40–45 kA in northern Florida. Apparently, this is the first report of an important lightning parameter, the return stroke current, varying over a large area of the globe. It seems that the generation of lightning strokes and their characteristics may be a sensitive function of latitude and hence of temperature. The results presented here indicate that even higher peak currents will be observed if this research is extended into the equatorial regions.

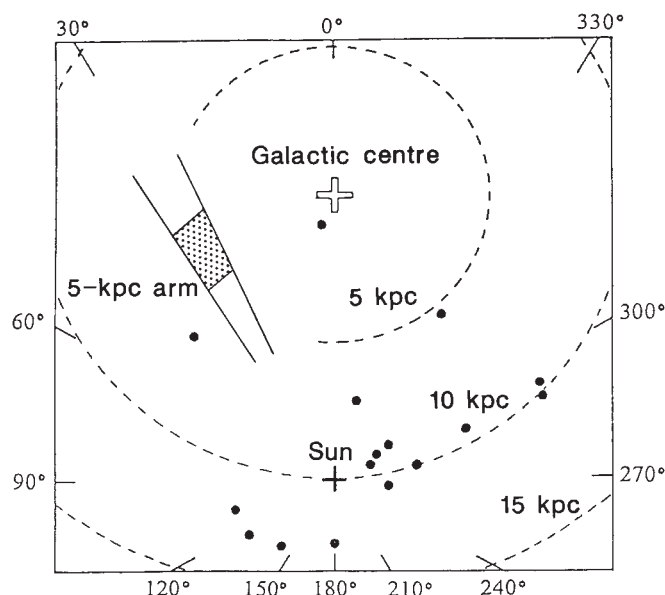


FIG. 2 The galactic distribution of catalogued X-ray pulsars¹. The most likely distribution of the new X-ray sources is shown by the dotted region.

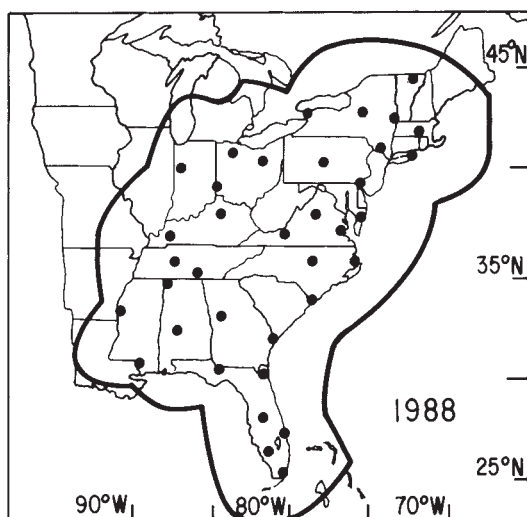


FIG. 1 The locations of 36 magnetic-direction-finder sites in the eastern United States are shown. The dark, irregular outline defines the area within which 70% of the cloud-to-ground lightning is detected by at least two direction finders. The irregular outline is produced in the following way. Circles with radii of 400 km, representing the nominal range of each direction finder, are drawn around each site. Where two or more circles overlap, a dark line is drawn to enclose the area of 70% detection efficiency. Outside this area, lightning flashes are detected, but with a lower efficiency.

The first three direction finders of the SUNY (State University of New York) Albany Lightning Detection Network were installed in New York in 1982 (ref. 6). The network was extended to Florida in 1985 and west to the Mississippi River in 1986. A full year of data was obtained in 1988 from the 36 direction finders (shown in Fig. 1). At each direction-finder site there is an orthogonal magnetic-loop antenna, a flat-plate antenna and associated electronics to process the incoming signals. The bandwidths of the antenna systems are $\sim 1\text{--}350$ kHz, so the shapes and polarities of the lightning waveforms are preserved. When the orthogonal magnetic-loop antenna senses the magnetic field from a lightning return stroke, a voltage is induced in each loop that is proportional to the product of the time derivative of the magnetic field and the cosine of the angle between the plane of the loop and the direction of propagation of the incoming field. The ratio of the integrated voltages in the orthogonal loops provides the direction of the lightning stroke. Processing electronics in the network are designed to respond only to those waveforms that are characteristic of return strokes in cloud-to-

ground flashes. Although some intracloud flashes may be accepted by the electronics, seven years of studies using video cameras have failed to identify even one intracloud flash among the several thousand flashes that were accepted by the network and videotaped during the studies.

All direction finders are controlled by microcomputers in the operations centre. Communication with the direction-finder sites is maintained by satellite at 9,600 baud. When a direction finder senses a lightning flash, the signals of up to 20 return strokes are processed and the time, bearing angle, signal amplitude and polarity of each stroke are transmitted to the operations centre. If two or more direction finders detect the flash within a programmed time interval, typically 6 ms, the optimum flash location is plotted⁷. Data, including the peak magnetic radiation field, are stored on tapes and optical disks for later analysis.

During 1988, the peak magnetic radiation field was recorded for the first return strokes of 5.3 million flashes that occurred in a range of 300 km from at least two of the sites shown in Fig. 1. Flashes at distances greater than 300 km were recorded, but I have excluded them from this analysis to reduce the bias of strong lightning flashes at large distances. This bias exists because the magnetic radiation field decays inversely with distance from the lightning flash; thus, the distribution of peak radiation values favours large currents at large distances ($\geq 300\text{--}400$ km). Here the bias is of no consequence because the significant observations are confined to the interior of the network where the typical direction-finder separation is of the order of 200 km.

Peak currents can be obtained from the peak radiation fields by using the results of an independent calibration that has been completed recently at the NASA Kennedy Space Center. Between 1985 and 1987, 17 triggered lightning return strokes were detected by the SUNY Albany Lightning Detection Network and their peak currents estimated using the theory of Uman *et al.*⁵. These estimates were then compared with the peak currents that were directly measured^{8,9} for the same 17 return strokes. A plot of the network peak-current estimates against the peak-current measurements yields a straight line of slope 0.87 and intercept 2 kA, and the correlation of the fit is 0.89. These results indicate the extent to which the network is providing realistic and reliable estimates of peak currents in lightning return strokes.

Figure 2 shows the geographical distribution derived for the mean peak currents of more than five million lightning first return strokes of both polarities in the eastern United States during 1988. I obtained these results by calculating the mean peak current at each grid point on a grid of 30×50 points (horizontal $\times$ vertical) and then forming the contours from the 1,500 grid values. Some large-scale variations are apparent in

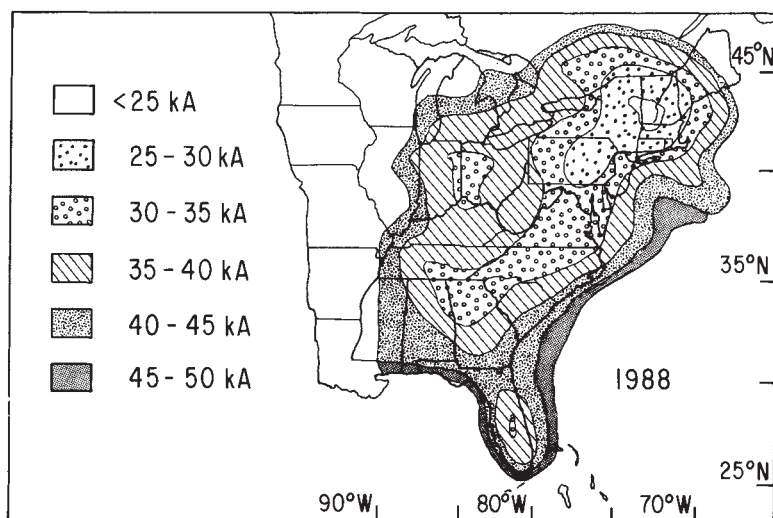


FIG. 2 Contours for the year 1988, which show the distribution by mean peak current of first return strokes to ground in lightning flashes of both polarities. I have used a database representing 5.3 million flashes.

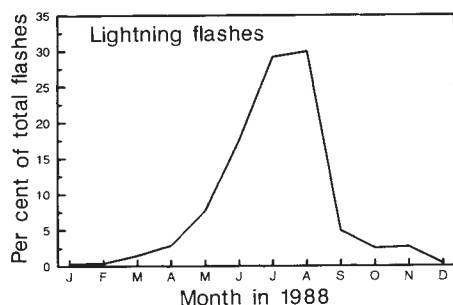


FIG. 3 Distribution by month of the flashes in the database used to produce Fig. 2, showing the predominance of summer lightning. There is some variation of the distribution with latitude, but the summer lightning dominates in the geographical areas I consider here.

the distribution of yearly mean peak currents of lightning first return strokes over the eastern United States. A minimum of ~ 25 kA occurs in New England, increasing monotonically along the east coast to the northern part of Florida. A region of lower peak currents is observed in central Florida. The variation with longitude of mean peak currents indicates that the lower values occur over the Appalachian Mountains. These observations of peak currents are confined to the interior of the network. Along the fringes of the network (see Fig. 1), higher peak currents are indicated. This is the effect of the bias, discussed previously, that reduces the number of weak lightning flashes detected as their distance increases from the direction finders. Coefficients of variation, that is, the ratio of the standard deviation to the mean, are ~ 0.5 in this data set.

The mean peak current may vary as a function of latitude because of the increasing volume of clouds associated with lightning at lower latitudes. Berger¹⁰ has shown from his measurements of 89 negative first return strokes to the instrumented towers on Mount San Salvatore (near Lugano, Switzerland) that the peak current I in a lightning return stroke can be related to the charge Q transferred to the ground. His empirical expression is $I = 10.6 Q^{0.7}$ where I is measured in kA and Q in coulombs. From this expression, a peak current of 25 kA (New England) corresponds to a total charge, assumed to reside on the channel, of 3.4 C; a peak current of 45 kA (Florida) corresponds to a total charge of 7.9 C. The ratio of 7.9 to 3.4 indicates that 2.3 times as much charge is involved in lightning first strokes in northern Florida as in New England. Cumulonimbus clouds are the source of the charge transferred in lightning and I discuss their characteristics below.

Figure 3 shows that most of the lightning (77%) in 1988 occurred in the summer months of June (18%), July (29%) and August (30%). Therefore, the change with latitude of such properties of summer cumulonimbus clouds as volume and temperature may explain the observed variation of the peak current. First, I consider the possible cloud volume variation with latitude by examining the cloud-top heights (measured using radar) on thunderstorm days for the summer of 1988. The analysis shows that in New England the cloud-top height averaged 12.5 km for 16 days, and in northern Florida it was 16.2 km for 20 days. I assume a summer cloud-base height of 1 km in both geographical regions, an approximation based on the available meteorological data. The cloud depth (base to top) is, therefore, 15.2 km in northern Florida and 11.5 km in New England and so the clouds in Florida have a volume larger than the clouds in New England by the ratio of the cloud depths, 15.2/11.5, or 1.3. Thus, there is 30% more volume available for the storage of electrical charge in the cumulonimbus clouds of Florida.

The relatively longer lightning channels that may occur in Florida could make another contribution to higher mean peak currents there. The centre for the negatively-charged region in

a cumulonimbus cloud occurs at a temperature level of $\sim -10^\circ\text{C}$ (ref. 11). Tables by Crutcher¹² show that in the summer the -10°C level is at 5 km over New England (45°N , -70°W) and 6 km over Florida (30°N , -80°W). This could indicate an increase of 20% in altitude of the lightning channel top, which would consequently produce an increase in the total charge on the channel available for the return stroke. The combination of an increase of cloud depth (30%) and channel length (20%) can, by the present argument, only account for about one-half of the observed increase in return-stroke peak current. If the cloud-base area, for example, in Florida thunderstorms is larger than in New England thunderstorms, the resulting extra cloud volume could store more charge. Although only a hypothesis so far, it seems that the greater volume of Florida cumulonimbus clouds and the longer lightning channels in Florida might account for the observations of higher peak currents in return strokes at lower latitudes. To test this, detailed observations of lightning return strokes in selected storms must be combined with meteorological observations of the storm clouds themselves. □

Received 13 June; accepted 28 November 1989.

- Pierce, E. T. *J. appl. Met.* **9**, 194–195 (1970).
- Takeuti T., Nakano, M. & Nagatani, M. *J. met. Soc. Jap.* **53**, 360–361 (1975).
- Harris, D. J. & Salzman, Y. E. *J. atmos. terr. Phys.* **34**, 775–786 (1972).
- Thomson, E. M. *J. geophys. Res.* **85**, 1050–1056 (1980).
- Uman, M. A., McLain, D. K. & Krider, E. P. *Am. J. Phys.* **43**, 33–38 (1975).
- Orville, R. E., Henderson, R. W. & Bosart, L. F. *Bull. Am. met. Soc.* **64**, 1029–1037 (1983).
- Orville, R. E. *J. geophys. Res.* **92**, 10877–10886 (1987).
- Fieux, R. P., Gary, C. & Hubert, P. *Nature* **257**, 212–214 (1975).
- Idone, V. P., Orville, R. E., Hubert, P., Barret, L. & Eybert-Berard, A. *J. geophys. Res.* **89**, 1385–1394 (1984).
- Berger, K. *Bull. Schweiz. electrotech. Ver.* **63**, 1403–1422 (1972).
- Williams, E. R. *J. geophys. Res.* **90**, 6013–6025 (1985).
- National Climatic Center *Selected Meridional Cross Sections of Heights, Temperatures and Dew Points of the Northern Hemisphere* NAVAIR-50-1C-59 (National Climatic Center, Asheville, North Carolina, 1970).

ACKNOWLEDGEMENTS. I thank R. W. Henderson and R. B. Pyle for their dedication to the operation of the SUNY Albany Lightning Detection Network and P. Moore for assistance in acquiring the meteorological data. This research was supported in part by the Electric Power Research Institute, by the NSF and by NASA.

Removal of nitrogen and sulphur oxides from waste gas using a phosphorus/alkali emulsion

S. G. Chang & D. K. Liu

Applied Science Division, Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720, USA

SULPHUR and nitrogen oxides emitted by the combustion of fossil fuels in power plants have an important role in acid-rain production¹. Most of the flue-gas desulphurization scrubbers installed to date involve wet limestone processes, which are efficient in SO_2 control but are incapable of removing nitric oxide. Here we report a new and efficient method for the combined removal of SO_2 and NO_x from flue gas, using aqueous emulsions containing yellow phosphorus and an alkali. The products comprise sulphate, nitrate and phosphate salts which can be used as fertilizer materials. The process described here may also be applied to stationary sources other than power plants, such as smelters, nitric acid plants and municipal incinerators for NO_x control.

The role of sulphur and nitrogen oxides in acid-rain formation and the destruction of lakes and forest ecosystems has been established^{2,3}. More stringent regulations of these emissions from both mobile and stationary sources are therefore imminent. The development of efficient processes for simultaneous SO_2 and NO_x control for power-plant flue gas is particularly important because fossil-fuel-fired steam boilers represent a major

source of sulphur and nitrogen oxide emissions. Existing wet flue-gas-desulphurization (FGD) scrubbers in power plants are incapable of removing flue-gas NO because of its low solubility in water. Several methods have been developed to enhance NO absorption in these scrubbers, including the use of oxidants to oxidize NO to the more soluble NO_2 (ref. 4) and the addition of various iron (II) chelates to bind and activate NO (refs 5–10). So far, none of these methods has been demonstrated to be cost effective despite high removal efficiencies of both SO_2 and NO_x .

We studied the removal of NO from flue gas by aqueous emulsions of yellow phosphorus (P_4) using a bench-scale gas scrubber. In a typical experiment, 1.0 g of yellow phosphorus was melted in 0.2 l of water at 60 °C in a Pyrex reaction column (50-mm inner diameter $\times$ 210 mm). A fine P_4 dispersion was created when a simulated flue-gas stream (~ 500 p.p.m. NO, 4% O_2 and the balance N_2) was bubbled through the reactor at a flow rate of $\sim 0.9 \text{ l min}^{-1}$. The scrubbed flue gas was passed through a gas washing bottle containing 0.2M NaOH and the NO and NO_2 concentrations in the exit gas were monitored using a Thermolectron Model 14A chemiluminescent NO_x analyser. The compositions of the spent scrubbing liquor and the NaOH absorbing solution were determined using a Dionex 2101i ion chromatograph.

We found that flue-gas NO can be effectively scrubbed by an aqueous emulsion of yellow phosphorus as shown in Fig. 1. The presence of O_2 is essential for the NO removal, and the system is more effective in the presence of excess O_2 in flue gas. We attribute the gradual decrease of NO removal efficiencies at longer absorption times to the increasing acidity of the scrubbing liquor, a problem that can be eliminated by the addition of an alkali to the system (see Fig. 2). Ion chromatographic analyses of the spent scrubbing liquor and the NaOH absorbing solution showed that all the NO absorbed was converted to a mixture of nitrite (NO_2^-) and nitrate (NO_3^-), and that $>95\%$ of the P_4 consumed can be recovered as a mixture of hypophosphite (H_2PO_2^-), phosphite (H_2PO_3^-) and phosphate (H_2PO_4^-).

NO and SO_2 can be simultaneously removed from flue gas by adding limestone (CaCO_3) to the phosphorus emulsion, as shown in Fig. 2. For this experiment, we used a ~ 1.2 -l jacketed reactor containing 0.9 l of a 0.3% w/w P_4 /5.0% w/w CaCO_3 slurry. The simulated flue gas was composed of 560 p.p.m. NO, 2,900 p.p.m. SO_2 , 10% O_2 and the balance was N_2 . In this case,

up to 100% of the flue-gas NO and SO_2 can be removed. Using laser Raman spectroscopy, the solid collected from the scrubber after the reaction was found to contain gypsum ($\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$), CaCO_3 and P_4 . The liquid phase in both the scrubber and the absorber contains (determined by ion chromatography) SO_3^{2-} , SO_4^{2-} , NO_2^- , NO_3^- , H_2PO_2^- , H_2PO_3^- and H_2PO_4^- , as shown in Fig. 3. In addition, the nitrogen-sulphur compounds hydroxy-imidodisulphate (HIDS) and imidodisulphate (IDS) are produced, as confirmed by ion chromatography¹¹. Both HIDS and IDS were subsequently hydrolysed to NH_4^+ when the scrubbing liquor was acidified to $\text{pH} < 2$. The formation of nitrogen-sulphur compounds by the reaction between NO_2^- and HSO_3^- in scrubbing liquor and their hydrolysis reactions have been reported previously^{5,12}. Therefore, the use of yellow phosphorus for combined NO and SO_2 removal can result in the production of sulphate, nitrate and phosphate salts of calcium or ammonium, all of which are valuable chemicals for the manufacture of fertilizers¹³.

The effectiveness for NO removal by yellow phosphorus was measured in a closed reaction system. A 1.0-l round-bottom flask containing 0.25 mg of P_4 , 0.1 g of CaCO_3 and 60 ml of deionized H_2O was evacuated and refilled with 1 atm of a gas mixture consisting of 500 p.p.m. NO, 10% O_2 and the balance N_2 . The reaction mixture was stirred magnetically at 50 °C until all P_4 had been consumed (~ 4 h). The composition of the reaction mixture was periodically analysed using ion chromatography to determine the NO removal effectiveness as defined by

$$\text{P/NO} = \frac{[\text{H}_2\text{PO}_2^-] + [\text{H}_2\text{PO}_3^-] + [\text{H}_2\text{PO}_4^-]}{[\text{NO}_2^-] + [\text{NO}_3^-]}$$

Our results showed that a P/NO ratio of 0.5 was obtained throughout the four-hour period. Experiments without yellow phosphorus indicated that NO removal through oxidation by O_2 to form NO_2 accounted for $<10\%$ of the total NO removed when P_4 was present. When 0.92 mM HSO_3^- (corresponding to $\sim 3,000$ p.p.m. of flue-gas SO_2) was added to the above reaction mixture, we obtained a P/NO ratio of 0.8.

We propose the following mechanism for NO removal by yellow phosphorus. It is well known that phosphorus vapour

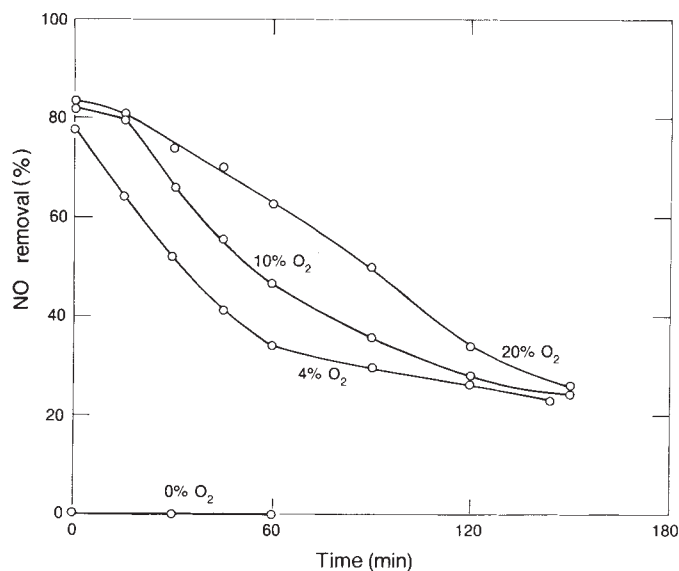


FIG. 1 NO removal by yellow-phosphorus emulsions as a function of O_2 partial pressure in flue gas. The reactions were carried out at 60 °C, and the initial and final pH were 3.5 and 1.5, respectively. The simulated flue gas contained 500 p.p.m. NO, 0–20% O_2 and the balance was N_2 .

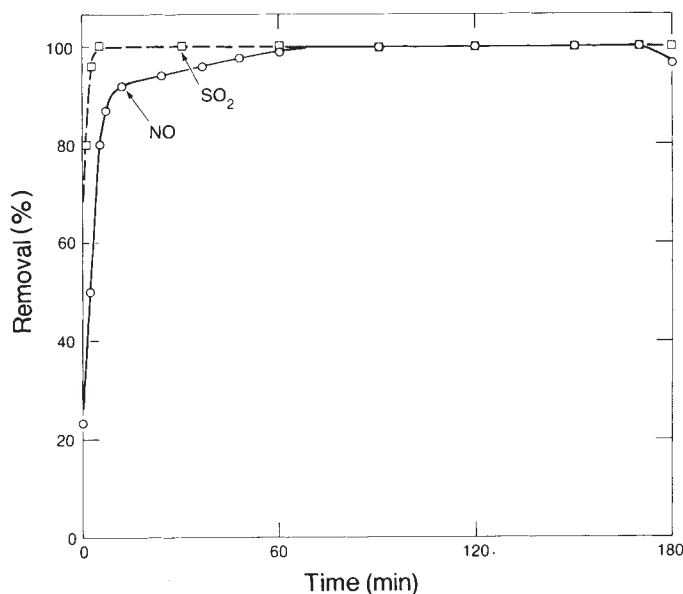


FIG. 2 Combined NO and SO_2 removal using an aqueous emulsion containing 0.8% w/w P_4 and 5.0% w/w CaCO_3 . The reaction was performed at 55 °C, and the initial and final pH were 7.5 and 4.2, respectively. The simulated flue gas contained 560 p.p.m. NO, 2,900 p.p.m. SO_2 , 20% O_2 and the balance was N_2 .

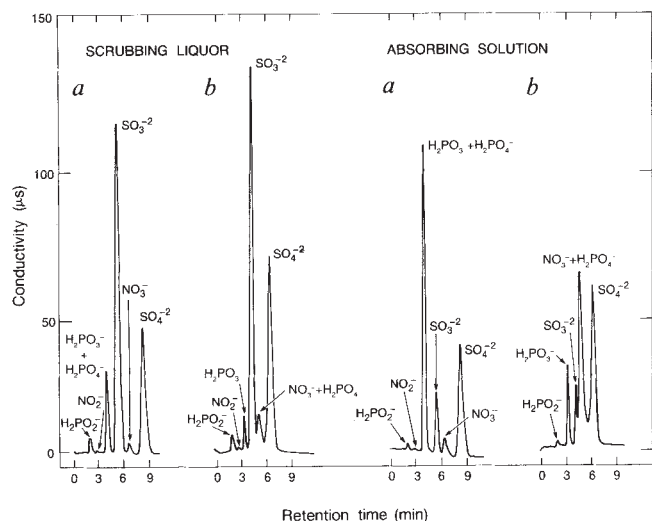


FIG. 3 Ion chromatograms of diluted (twenty-five times) spent scrubber and absorber solutions from the combined NO and SO₂ removal experiment with P₄ and CaCO₃. We used a Dionex 2101i ion chromatograph equipped with a conductivity detector and a Dionex AS4 anion separation column for the analyses. The eluants were: a, 5.3 mM Na₂CO₃ and b, 4.0 mM Na₂CO₃/2.0 mM NaOH/0.5% v/v CH₃CN.

reacts with oxygen gas in the presence of moisture to give various phosphorus oxides, atomic oxygen and ozone¹⁴⁻¹⁶. The O and O₃ thus produced can oxidize NO to NO₂, which can then react with another molecule of NO to form N₂O₃ or dimerize to form N₂O₄. Both N₂O₃ and N₂O₄ are much more soluble in water than NO, and the HNO₂ and HNO₃ produced by the hydrolysis of N₂O₃ and N₂O₄ are neutralized by alkali in the scrubbing liquor (such as CaCO₃) to yield NO₂⁻ and NO₃⁻. The mechanism for the oxidation of phosphorus to various phosphorus oxides seems to be complicated¹⁴⁻¹⁶ and is being investigated.

We have done a preliminary economic consideration of the above phosphorus-based process which suggests that the technology warrants further development. The addition of yellow phosphorus to a limestone FGD system to achieve simultaneous control of NO_x emissions does not require excessive further equipment because the processing of the scrubbing liquors seems to be rather straightforward. The recovery of phosphorus by-products can be achieved simply by precipitation at appropriate pH of the liquors. At the operating pH of a limestone scrubber (between 3.5 and 5.5), the majority of the phosphate is in the form of H₂PO₄⁻. The calcium salt Ca(H₂PO₄)₂ · H₂O is relatively water soluble under such conditions and can be separated from the less soluble CaSO₃ · ½H₂O and CaSO₄ · 2H₂O by simple filtration. Limestone can then be added to the filtrate to raise the pH to 8 or higher where H₂PO₄⁻ is converted to HPO₄²⁻. The calcium salt CaHPO₄ · 2H₂O is quite insoluble under such conditions and can be recovered as a precipitate. Therefore, much of the value of yellow phosphorus (\$0.91 per lb)¹⁷ consumed can be recovered as CaHPO₄ · 2H₂O, which is a valuable commercial product used as animal feed and fertilizer (\$0.63 per lb of P)¹⁷. At a P/NO ratio of unity, the loss in the value of phosphorus reagent (\$560 per ton of NO removed) is small compared to the cost of a selective catalytic reduction process¹⁸. With an improvement of P/NO ratio, this loss in chemical value can be further reduced. The soluble nitrogen-sulphur compounds formed in the system can be removed from the solutions either by precipitation as potassium salts or by acid hydrolysis^{5,12}. We are conducting a scale-up test of this phosphorus-based process on a larger bench-scale system (gas flow rate = 600 l min⁻¹) and a realistic economic evaluation of a commercial-plant design will be possible after integrated pilot-scale tests. □

Received 31 July; accepted 17 November 1989.

1. Wark, K. & Warner, C. F. *Air Pollution-Its Origin and Control* (Harper & Row New York, 1981).
2. Cosby, B. J. et al. *Environ. Sci. Technol.* **19**, 1144-1149 (1985).
3. Hileman, B. *Environ. Sci. Technol.* **18**, 8A-10A (1984).
4. Yaverbaum, L. H. *Nitrogen Oxides Control and Removal* (Noyes Data Corp., New Jersey, 1979).
5. Chang, S. G. in *Fossil Fuels Utilization: Environmental Concerns* (eds Markuszewski, R. & Blaustein, B. D.) 159-175 (American Chemical Soc., Washington, DC, 1986).
6. Liu, D. K., Frick, L. P. & Chang, S. G. *Environ. Sci. Technol.* **22**, 219-223 (1988).
7. Liu, D. K. & Chang, S. G. *Environ. Sci. Technol.* **22**, 1196-1200 (1988).
8. Chang, S. G., Littlejohn, D. & Liu, D. K. *Ind. Eng. Chem. Res.* **27**, 2156-2161 (1988).
9. Zhuang, Y. & Deng, Y. *Acta Sci. Circumstantiae* **5**, 347-353 (1985).
10. Bedell, S. A., Tsai, S. S. & Grinstead, R. R. *Ind. Eng. Chem. Res.* **27**, 2092-2095 (1988).
11. Littlejohn, D. & Chang, S. G. *Analyt. Chem.* **59**, 158-160 (1986).
12. Chang, S. G., Littlejohn, D. & Lin, N. H. in *Flue Gas Desulfurization* (eds Hudson, J. L. & Rochelle, G. T.) 127-152 (American Chemical Soc., Washington, DC, 1982).
13. Huffman, E. O. in *Kirk-Othmer Encyclopedia of Chemical Technology* Vol. 10, 3rd edn (eds Bushey, G. J. et al.) 31-125 (Interscience, New York, 1980).
14. Cordes, H. & Witschel, W. *Z. phys. Chem.* **46**, 35-48 (1965).
15. VanZee, R. J. & Khan, A. U. *J. chem. Phys.* **65**, 1764-1772 (1976).
16. Andrews, L. & Withnall, R. *J. Am. chem. Soc.* **110**, 5605-5611 (1988).
17. Chemical Marketing Reporter, 7 November 1988, pp. 34-44.
18. Eskinazi, O. D. & Cichanowicz, J. E. *J. Air Pollution Control Ass.* **39**, 1131-1139 (1989).

ACKNOWLEDGEMENTS. This work was supported by the US Department of Energy through the Pittsburgh Energy Technology Center.

Palaeobotanical evidence for a marked temperature increase following the Cretaceous/Tertiary boundary

Jack A. Wolfe

US Geological Survey, MS-919, Box 25046, Federal Center, Denver, Colorado 80225-0046, USA

CORRESPONDENCE analysis of dicot leaf physiognomy of modern vegetational samples from a wide range of environments indicates that >70% of physiognomic variation corresponds to water or temperature factors, or both. Despite wide variation in single physiognomic characters, overall trends can be used to distinguish between samples from different climates. Some climate parameters are well correlated with changes in physiognomy, so that climate characteristics can be inferred from physiognomic analyses. Here I apply this climate-leaf analysis multivariate program (CLAMP) to leaf assemblages from the Cretaceous/Tertiary boundary. The results indicate a fourfold increase in precipitation at the boundary and an increase in mean annual temperature of 10 °C. These levels persisted for 0.5-1.0 Myr, after which precipitation decreased to about three times the values for the latest Cretaceous, and the mean annual temperature decreased to 5-6 °C above latest Cretaceous values.

The analysis of morphological and anatomical adaptations (physiognomy) of land plants to environment is a direct and common method of inferring ancient land climates¹⁻⁴. The aerial vegetative parts (woods and leaves) of plants adapt in varying degrees to environment^{3,5}, and leaf physiognomic characters have a physiological basis⁶. Previous physiognomic analyses of fossil land-plant assemblages used univariate correlations of particular physiognomic categories with particular environmental parameters (such as the ratio between the percentage of entire-margined (untoothed) dicot species and the mean annual temperature (MAT)), but such correlations may be imprecise⁷. Characters of dicot leaves other than margin type respond to environment, but the relation of these characters to climate is known only in general and/or relative terms, and the possible interrelations of characters are unknown. Further, the degree to which environmental correlation of any single physiognomic character might be altered by, for example, higher levels of carbon dioxide, is unknown.

Despite such uncertainties, physiognomic analyses have provided some indications of the change in the environment across

TABLE 1 Data for evaluating significance of climatic variables

Variable	Eigenvalue	Axis I	Eigenvalue	Axis II
		Total variance (%)		Total variance (%)
None (physiognomy only)	0.224	52.5	0.088	20.7
Mean annual temperature	0.215	52.3	0.086	21.0
Mean temperature, warm month	0.212	52.4	0.085	20.9
Mean temperature, cold month	0.219	52.2	0.089	21.3
Mean annual precipitation	0.228	53.1	0.089	20.7
Mean precipitation, 5 dry months	0.234	54.8	0.084	19.8
5 warm months, temperature/precipitation	0.251	56.8	0.084	19.1
5 dry months, temperature/precipitation	0.243	54.9	0.086	19.5

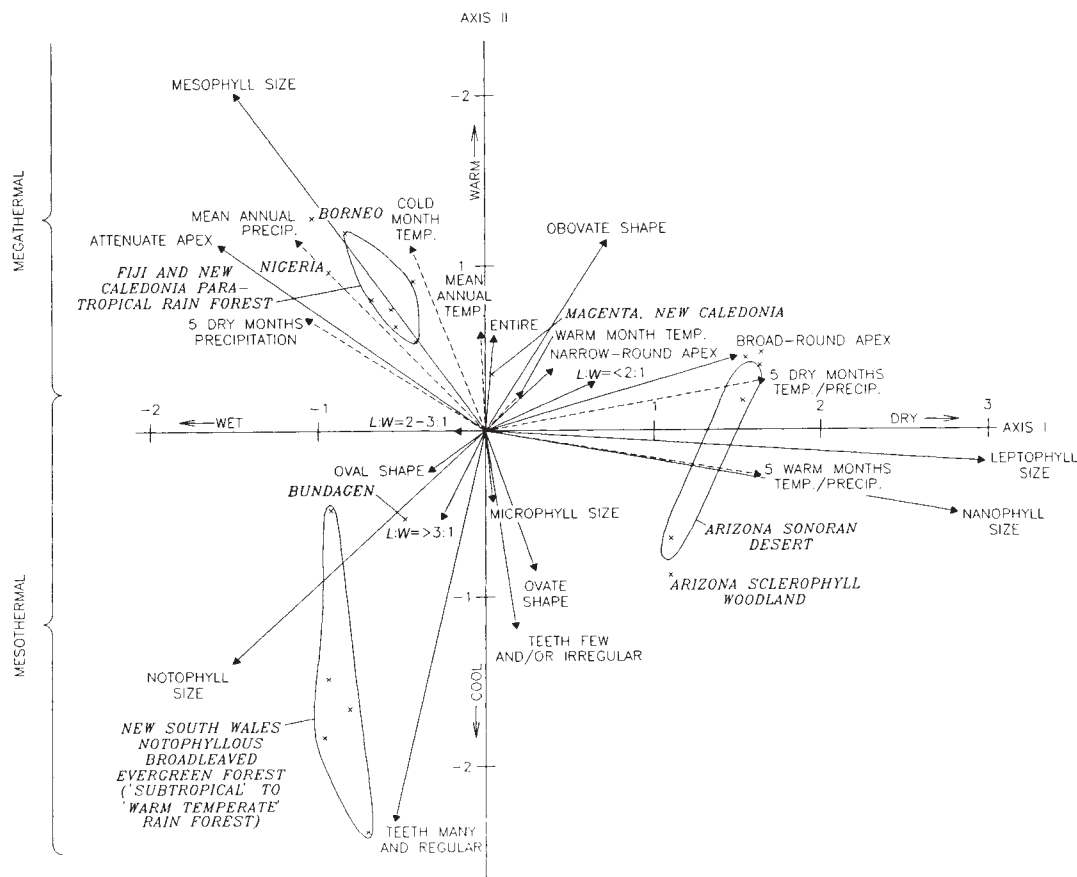
The high percent variance on Axis II for temperature variables indicates that Axis II corresponds to relative warmth, and the higher percent variance on Axis I for precipitation variables, especially for those that are related to evapotranspiration, indicates that Axis I corresponds to water stress.

the Cretaceous/Tertiary (K/T) boundary in the western interior of North America. Using only the entire-margin correlation, a gradual temperature decrease across the K/T boundary was suggested⁴. Later analyses¹⁻³ based on three univariate correlations indicated: (1) a major, unspecified precipitation increase; (2) no overall temperature change across the boundary, although indirect evidence indicated a 1-2 month low-temperature excursion; (3) that, to explain the divergence from the earlier analysis, some physiognomic characters could be out of phase with the environment because of historical factors such as a low-temperature excursion after a bolide impact.

To obtain greater validity (accuracy) and reproducibility (pre-

cision) in physiognomic analyses, I collected samples of modern plants from sites with climates having a variety of temperatures, precipitation and annual variation. Physiognomic characters that are readily determinable from fossil-leaf impressions were tabulated for each species in a sample; characters of minor, if any, taxonomic significance were assumed to be environmentally dependent. The percentages of species having each physiognomic character state in a sample were subjected to correspondence analysis (reciprocal averaging^{8,9}), which provides a method of inferring the patterns of correlation in the data set and of positioning the physiognomic characters and samples on two (or more) axes.

FIG. 1 Reciprocal averaging plots for samples of modern vegetation, leaf-physiognomy character states (solid lines), and climate variables (dashed lines). Moist megathermal (MAT > 20 °C) samples plot in the upper left quadrant, moist mesothermal (MAT 13–20 °C) samples in the lower left quadrant, and subhumid to arid samples in the right quadrants. The Magenta sample represents largely open-canopy, low (7–8 m high) woodland. The Bundagen sample represents a partially open-canopy, low (8–10 m) forest that grows on sandy soils. Physiognomic vectors are extended to their score points and represent characters that distinguish samples: for example, megathermal rain forests in the upper left quadrant are characterized by large leaves that have attenuate apices (drip tips), entire margins (shared with dry megathermal vegetation), and a length:width (L:W) ratio of 2–3:1 (shared with moist mesothermal vegetation). Each climate variable was derived from meteorological



data for each sample; each climate variable was entered into the physiognomic data base and run separately, as if a climate variable were a physiognomic character. For minimum distortion of the character plots, the units of measurement selected for each variable have approximately the same ranges when expressed as percentages, and the ratios were converted to percentages by letting the maximum ratio be 100. Mean annual temperature

and entire margin are both almost coincidental with Axis II, suggesting that the entire margin character is generally the best single character for estimating MAT. The physiognomic data base is a tabulation from collections made in 1988, except for the samples from Nigeria and Borneo, which were tabulated from ref. 5.

TABLE 2 Temperature and precipitation for K/T-boundary leaf assemblages

Time interval	Assemblage	I and II		Temperature		Precipitation	
		Number of species*	Total variance (%)	CLAMP MAT (°C)	Latitude MAT† (°C)	CLAMP MAP‡ (cm)	CLAMP 5 dry (cm)
Phase 5	Genesee	22	69.0	17.7	18.1	160	50
	Clareton	11	70.2	19.8	21.3	250	80
	South Table Mountain	22	72.2	22.8	22.7	210	70
	Animas	45?	72.1	23.6	23.4	230	70
	Raton	28	73.0	23.5	23.5	260	80
Upper Phase 4	Raton	11	72.0	>27.4	—	>350	>100
Phase 4	Raton	19	72.6	>27.4	—	>350	>100
Phase 3/lwr 4	Brownie Butte	9	67.8	22.8	>23.9	>350	>100
	Raton	10	72.2	>27.4	>27.4	>350	>100
Phase 3	Raton	3	68.0	>27.4	—	>350	>100
Maastrichtian (late)	Lance	29	71.6	16.2	16.2	70–80	25
	Medicine Bow	37	72.0	16.3	16.5	80–90	25–30
	Laramie	82?	71.6	18.6	17.2	80–90	25–30
	Littleton	36	72.4	17.3	17.3	90–100	30–35
	Rockdale	30	70.3	17.8	17.6	80–90	25–30
	Raton Phase 1	21	70.7	18.1	18.2	80–90	25–30
	Vermejo	32	69.6	18.1	18.2	80–90	25–30
	Ripley	79?	72.4	16.8	—	70–80	25
(early)							

* The number of species in some assemblages (denoted by ?) is based on outdated taxonomy. In some instances, the number of species cited here is different from ref. 1 because of continuing taxonomic work.

† Latitude MAT estimates are calculated by assuming the validity of one CLAMP MAT estimate (the assemblage with the highest total variance,) and a 'best fit' latitudinal temperature gradient of 0.33 °C per 1° latitude, which is the same value suggested on other grounds by ref. 2.

‡ For the Maastrichtian estimate of mean annual precipitation (MAP), I assume that precipitation was almost aseasonal (as indicated by fossil woods²), and thus the estimate of MAP is derived from the estimate of precipitation for the five dry months, which can be more validity inferred from CLAMP for assemblages that plot in the subhumid area.

In the resulting plot (Fig. 1), Axis I is related to water-stress adaptations, with Sonoran Desert samples positioned to the far right and rain-forest samples to the left. Water stress, as measured by ratios of warm- and/or dry-season temperature to precipitation (Table 1), closely corresponds to Axis I. Axis II corresponds to warmth: the values of megathermal samples lie above and those of mesothermal samples below (microthermal samples are not plotted here but fall still lower). On Axis II, megathermal samples are compressed about three times as mesothermal samples but the placement corresponds to MAT. When allowance is made for this compression, a regression analysis with MAT as the variable produced a squared correlation coefficient, r^2 , of 0.71, with a 0.6 °C standard error of MAT.

The physiognomy of a given sample or leaf assemblage represents (ignoring taphonomic biases of fossil assemblages) complex compromises between different environmental variables; values of any given character in different samples may be determined by such compromises. For example, one Fiji sample has 55% mesophylls and 32% attenuate apices, and a second has 36% mesophylls and 57% attenuate apices; both samples came from low altitude and similar climate and their values lie close together on the plot. If leaf size or apex were used alone, one might conclude that these samples had grown under different climates. The application of CLAMP to K/T-boundary assemblages allows more accurate and precise evaluations of the directions and magnitudes of changes in climatic variables during this interval than do single-character correlations.

The Vermejo and Raton Formations (Fig. 2b) contain the most complete and most studied sequence of K/T-boundary leaf assemblages³. Assemblages were collected from <30 cm below the horizon that contains iridium enrichment and shock-metamorphosed minerals, which are strong evidence of bolide impact^{10,11}. Phase 2 (<1 m above this horizon; corresponds partly to the palynological fern-spore spike¹²) contains only ferns and so CLAMP cannot be used. Phase 3 (beginning <1 m above Phase 2) contains depauperate leaf assemblages; numerous, more diverse assemblages occur in Phases 4 and 5. The boundary between 4 and 5 is estimated to be ~0.5–1.0 Myr after the impact³. Other western interior sequences also yielded leaf assemblages that are stratigraphically related to the impact

horizon or the K/T boundary determined from vertebrate, radio-metric or palaeomagnetic data¹.

Placement (Fig. 2a) of late Maastrichtian samples relative to early Paleocene samples indicates a fourfold precipitation increase at the K/T boundary (Table 2). This humid climate prevailed for 0.5–1.0 Myr, and then the precipitation decreased to levels about three times the Maastrichtian level. A minor (~10 cm) precipitation increase from early late to latest Maastrichtian could be inferred in the Denver (but not in the Raton) Basin and may correlate with an increase inferred from soil change in North Dakota¹³.

A major 10 °C MAT increase at the K/T boundary is also inferred from the CLAMP results. In previous MAT estimates^{1–3} for late Maastrichtian leaf assemblages, the present Southern Hemisphere ratio between entire-margined species and MAT has been assumed to be applicable to late Cretaceous assemblages, and thus the MAT for both late Maastrichtian and early Paleocene assemblages in the Raton and Denver Basins was estimated to be in the cool megathermal range. CLAMP results, however, indicate that the MAT increased significantly at the boundary, remained high for 0.5–1.0 Myr, and then decreased, but it was still well above the late Maastrichtian MAT. The MAT might have gradually decreased from a high shortly after the boundary; however, when Phase 3 is combined with the lower part of Phase 4 and when the upper part of Phase 4 is considered alone, the runs have sample scores very similar to those for Phase 4 (Fig. 2a).

Climate changes at the K/T boundary altered, at least briefly, some relations between physiognomic characters and environment, such as the ratio between margin and MAT. The Brownie Butte assemblage has 27% entire-margined species¹, indicating a MAT of ~10 °C (ref. 4), compared to this CLAMP estimate of 23 °C which is consistent with other CLAMP estimates for the same period (Table 2). Some characters (such as size and apex) change during life of a plant depending on environment^{5,6}, and the magnitude of these changes was apparently sufficiently great at the K/T boundary to override, in a multivariate analysis, the alteration in margin types.

Despite evidence for major warming across the K/T boundary, a low-temperature excursion seems to have occurred before the warming¹. Selective survival of plants that could enter

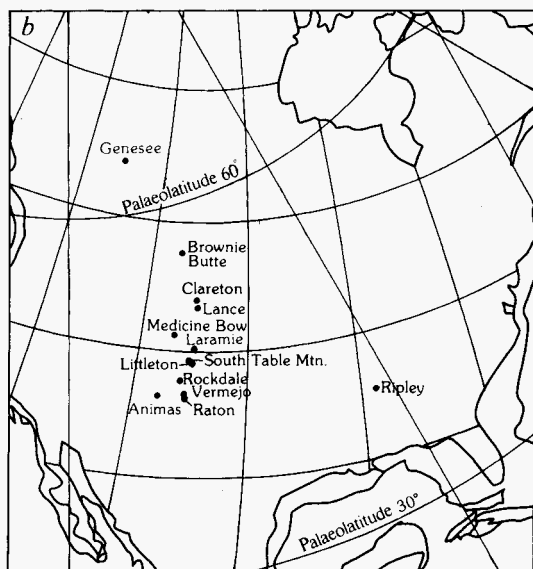
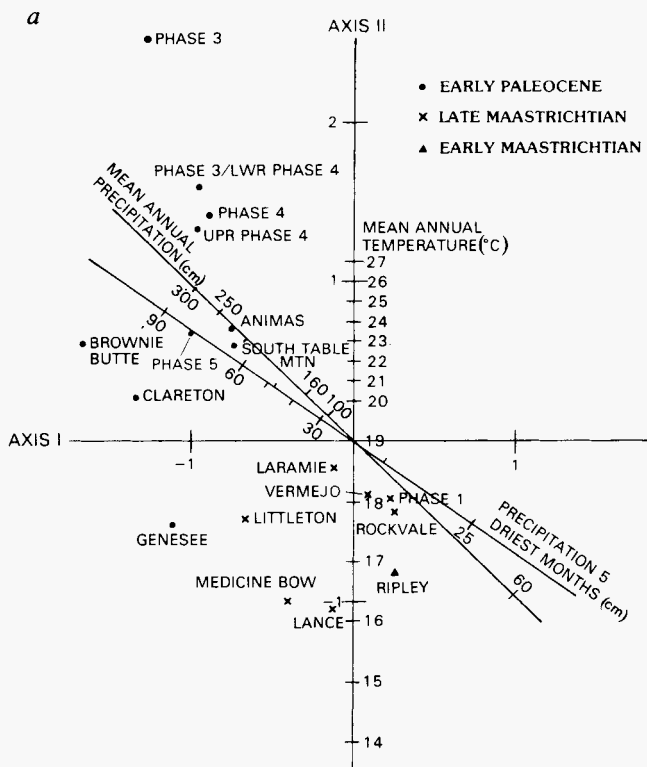


FIG. 2 Leaf assemblages from the K/T boundary showing *a*, the reciprocal averaging plot, and *b*, geographic positions. Comparisons of palaeolatitudes with inferred palaeotemperatures indicate the high degree of internal consistency in sets of CLAMP temperature estimates. Each assemblage was entered separately into CLAMP data base and run separately. None of the runs significantly altered the modern sample scores, except for the small assemblage of Raton Phase 3, which resulted in halving Axis II scores for moist megathermal assemblages. Results on Raton Phase 4, Raton Phase 3 and lower Phase 4, and upper Phase 4, slightly lowered scores for moist megathermal assemblages. The early Maastrichtian Ripley assemblage is from the Mississippi Embayment and shows that MAT increased moderately from the early into the late Maastrichtian. Palaeolatitudes derived from data in ref. 17.

dormancy (deciduous dicots and conifers, ferns and herbaceous monocots) over evergreen plants indicates an event that favoured dormancy¹⁴, but warming, accompanied by a major precipitation increase, would not favour dormancy.

A major temperature increase shortly after the K/T boundary is consistent with inferred greenhouse heating^{15,16}. An oceanic impact site, resulting in the injection of large amounts of water vapour into the stratosphere and the formation of a humid greenhouse is suggested in one model¹⁵; however, the stratospheric residence time for this water vapour would be of the order of months or years. Because warmth and wetness continued for a far longer time, complex feedback mechanisms in the earth's ocean-atmosphere system may have altered the carbon cycle² and may have involved factors such as production of large amounts of carbon dioxide by the bolide impact¹⁶. □

Received 3 April; accepted 13 November 1989.

1. Wolfe, J. A. & Upchurch, G. R. *Nature* **324**, 148–152 (1986).
2. Wolfe, J. A. & Upchurch, G. R. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **61**, 33–77 (1987).
3. Wolfe, J. A. & Upchurch, G. R. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5096–5100 (1987).
4. Hickey, L. J. *Nature* **292**, 529–531 (1981).
5. Richards, P. W. *The Tropical Rain Forest* (Cambridge University Press, 1952).
6. Parkhurst, D. F. & Loucks, O. L. *J. Ecol.* **60**, 505–537 (1972).
7. Greenwood, D. R. thesis, Univ. Adelaide (1988).
8. Hill, M. O. *J. Ecol.* **61**, 237–249 (1973).
9. Hill, M. O. *Appl. Statistics* **23**, 340–354 (1979).
10. Pillmore, C. L., Tschudy, R. H., Orth, C. J., Gilmore, J. S. & Knight, J. D. *Science* **223**, 1180–1183 (1984).
11. Bohor, B. F., Foord, E. E., Modreski, P. J. & Triplehorn, D. M. *Science* **224**, 867–869 (1984).
12. Tschudy, R. H., Pillmore, C. L., Orth, C. J., Gilmore, J. S. & Knight, J. D. *Science* **225**, 1030–1032 (1984).
13. Fastovsky, D. E. & McSweeney, K. *Geol. Soc. Am. Bull.* **99**, 66–77 (1987).
14. Upchurch, G. R. in *Mass Extinctions: Processes and Evidence* (ed. Donovan, S. K.) 195–216 (Columbia University Press, 1989).
15. Emiliani, C., Kraus, E. B. & Shoemaker, E. M. *Earth planet. Sci. Lett.* **55**, 317–334 (1981).
16. O'Keefe, J. D. & Ahrens, T. J. *Nature* **338**, 247–249 (1989).
17. Smith, A. G., Hurley, A. M. & Briden, J. C. *Phanerozoic Paleogeographic World Maps* (Cambridge University Press, 1981).

ACKNOWLEDGEMENTS. Modern samples were collected with the help of A. Tuisawau and other personnel of the University of the South Pacific, T. Jaffré (ORSTOM in Noumea), A. Floyd (New South Wales National Parks and Wildlife Service) and the personnel of the Forestry Commission of New South Wales. W. Kovach supplied the program for reciprocal averaging.

Palaeontological and isotope evidence for warm saline deep waters in Ordovician oceans

L. Bruce Railsback*§, Spafford C. Ackerly†||, Thomas F. Anderson* & John L. Cisne†‡

* Department of Geology, University of Illinois, 245 Natural History Building, 1301 West Green Street, Urbana, Illinois 61801–2999, USA

† Department of Geological Sciences and Institute for the Study of the Continents, Snee Hall and, ‡ Division of Biological Sciences, Cornell University, Ithaca, New York 14853, USA

TODAY'S oceans are characterized by cold bottom waters, produced at high latitudes. Here we present morphological and oxygen isotope data for brachiopods from the Caradocian (Upper Ordovician) Trenton Group, in New York, which indicate that both temperature and salinity increased with depth in Ordovician oceans. Production of warm, saline deep waters in low- to mid-latitude evaporative seas, and reduced production of cold deep water, may have been favoured by high sea level, warm global climate and fortuitous positioning of continents^{1–5}. The resulting ocean circulation system was markedly different from that of today. Lower oxygen contents of warm, saline deep waters may have contributed to the deposition of organic-rich sediments, which are abundant in Caradocian strata^{6,7} but rare in modern ocean sediments.

§ Present address: Department of Geology, University of Georgia, Athens, Georgia, 30602, USA.

|| Present address: Department of Geology, Carnegie Building, Oberlin College, Oberlin, Ohio, 44074–1044, USA.

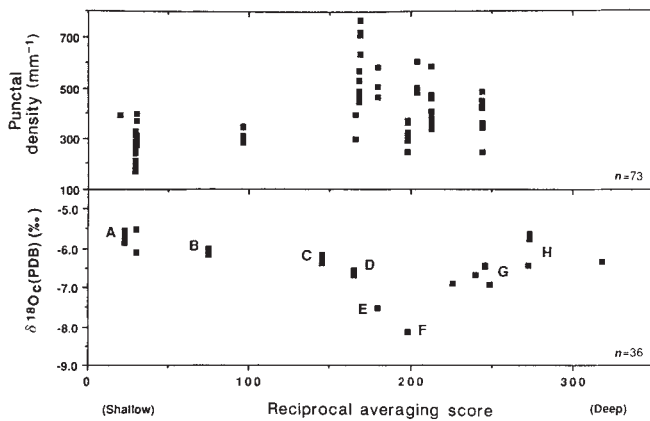


FIG. 1 Data from *Paucicrura rogata*, an orthid brachiopod in the Trenton Group of New York, along one time-parallel transect. *a*, Variation in density of punctae on shells with respect to reciprocal averaging score, a measure of palaeodepth⁹. *b*, Variation in oxygen isotopic composition of shells with palaeodepth. Letters near data clusters are used in Fig. 2 to locate samples in *T-S* space.

We sampled fossil brachiopods (calcitic bivalved marine bottom dwellers) from the Caradocian Trenton Group in the Mohawk Valley of central New York to obtain palaeontological and chemical evidence to test the hypothesis that Ordovician oceans were characterized by warm saline deep waters. Trenton Group strata in the Mohawk Valley were deposited on the eastern continental slope of North America during its abortive subduction by the encroaching Ammonoosuc volcanic arc, in a plate-tectonic setting similar to that of the modern Australian Shelf and Timor Trough⁸. This modern analogue, in addition to sedimentological and palaeontological considerations, indicates that the brachiopods lived over a depth range from upper slopes (with palaeodepths in tens of metres) down to 2–3 km (ref. 9). Circulation in the modern analogue and the palaeobiogeographic similarity of the Trenton Group fauna to that from other regions^{10,11} indicate that the Taconic Basin was in open communication with the global ocean. The basin's depth in a subductive trench should have been appropriate for estimation of ocean conditions.

Palaeodepth relations between individual samples (but not absolute palaeodepths) can be estimated from the ordination (Cisne *et al.*⁹) of the entire fossil assemblages in which the brachiopods occur. This ordination yields a reciprocal averaging score roughly proportional to palaeodepth for each field sample. This ecological measure of palaeodepth is more accurate than

simple geographic distance down the depositional slope because syn-sedimentary faulting generated horsts and grabens along the slope.

Paucicrura rogata is an orthid brachiopod that lived along much of the Trenton palaeoslope. Gradients in environmentally controlled morphological features and in oxygen isotopic compositions of shells of this species thus provide evidence (independent of interspecific biological effects) regarding marine temperature and salinity gradients. We examined cathodoluminescence and microtextures of the brachiopods where possible to avoid sampling of shells altered after deposition. We used only samples with very minor or no cathodoluminescence for palaeoceanographic considerations.

We also excluded¹² samples with low ¹³C content (an apparent indicator of alteration). The oxygen isotopic compositions of the brachiopods cannot be readily interpreted in terms of known burial diagenetic trends in the Mohawk Valley¹², which also indicates that the data reflect the depositional conditions. The general inverse relationship between punctal density (an anatomical, unalterable characteristic) and isotopic composition shown in Fig. 1 also indicates that the oxygen isotope data reflect depositional conditions.

Abundance of punctae (pores connecting the shell interior and exterior) in modern brachiopods is roughly proportional to the temperature of the water in which they live^{13,14}. Figure 1*a* shows the density of punctae in Trenton Group brachiopods plotted against reciprocal averaging score (that is, palaeodepth). Punctal densities increase with palaeodepth over the upper reaches of the palaeoslope, indicating that water temperatures increased with depth in Ordovician oceans over most of the range sampled.

Oxygen isotopic compositions ($\delta^{18}\text{O}(\text{‰}) = [({}^{18}\text{O}/{}^{16}\text{O})_{\text{sample}}/({}^{18}\text{O}/{}^{16}\text{O})_{\text{standard}}] - 1$, relative to PDB standard) of well preserved Trenton Group brachiopod shells decrease with depth over the upper part of the palaeoslope, but increase again at great depth (Fig. 1*b*). The decreasing $\delta^{18}\text{O}$ values indicate either increasing temperatures (and thus less fractionation of ¹⁸O into brachiopod calcite) or drastically decreasing $\delta^{18}\text{O}$ of sea water with depth during deposition. The data are modelled in Fig. 2 using a graphical technique¹⁵ in which we constrain the temperature and salinity using the calcite (shell) isotopic compositions and the increasing seawater density. In this technique, samples must be located on their appropriate isopleths for isotopic composition and must be in a sequence in which the density increases with palaeodepth. Modelling of the Trenton Group brachiopod data indicates that temperature and salinity increased with depth (Fig. 2). The Ordovician near-surface or intermediate waters had relatively low temperatures (13–19 °C) and low salinities (26–29 parts per thousand (p.p.t.)). Deeper waters were warmer

FIG. 2 Modelling¹⁵ of temperature and salinity for samples shown in Fig. 1*b*. Dashed lines are lines of equal seawater density (isopycnals, lines of equal density, expressed as σ_t); density of sea water must increase with depth. Solid lines are lines of equal calcite isotopic composition (PDB standard). Lettered points represent data clusters in Fig. 1*b*, in alphabetical order for increasing palaeodepth. Locations of samples in *T-S* space are constrained by increasing seawater density with depth and by oxygen isotopic composition of brachiopod shells.

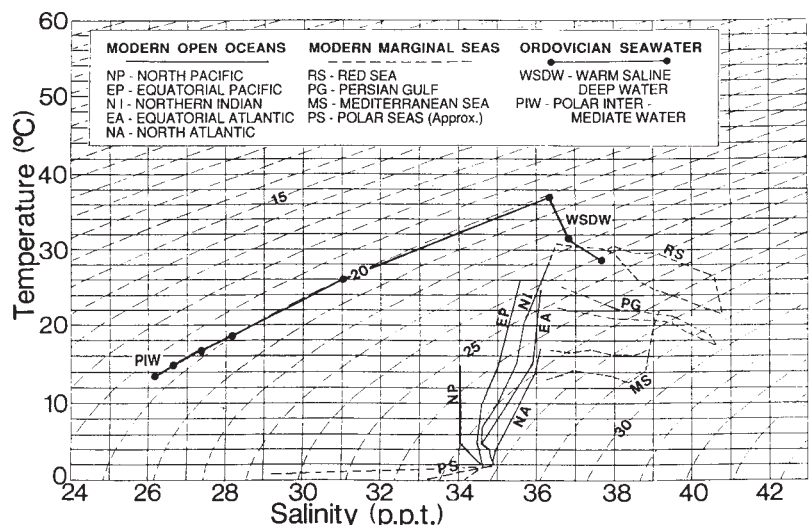
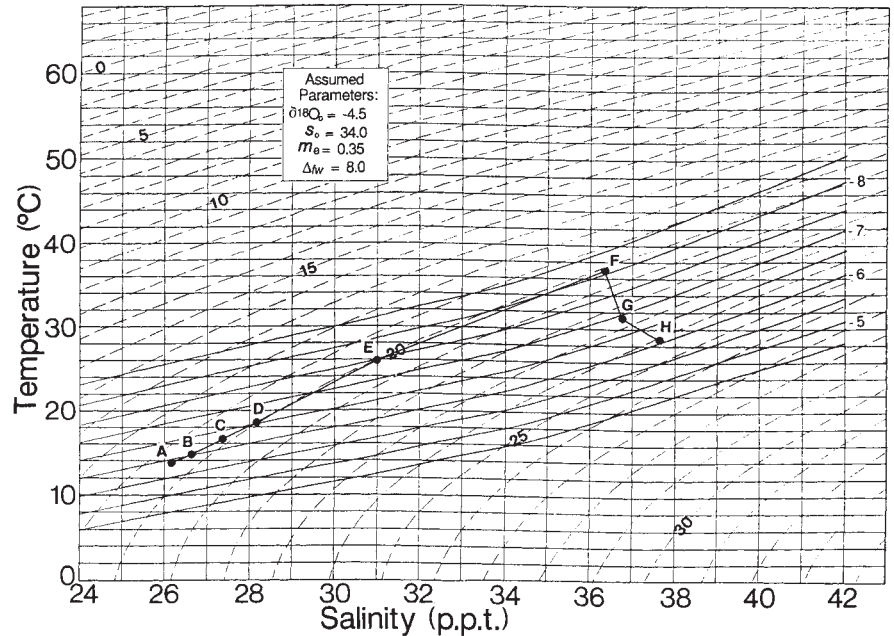


FIG. 3 Temperature-salinity profile implied by Ordovician samples (from Fig. 2, temperature and salinity increase with depth) plotted with T - S profiles in modern open oceans¹⁷ (where temperature decreases with depth) and T - S variation in modern marginal seas¹⁸⁻²¹. Dashed lines are isopycnals; deeper waters in any sequence have greater densities.



and more saline (28–38 °C and 36–38 p.p.t.) (Fig. 2). The precision of these results depends on assumptions about the average isotopic composition and average salinity of Ordovician sea water ($\delta^{18}\text{O}_0$ and S_0 in Fig. 2). Precision also depends on assumptions about the relationship between isotopic composition and salinity in waters more brackish (Δf_w in Fig. 2) or more saline (m_0) than the global average. However, the slope of the samples in temperature-salinity space in Fig. 2, depends little on those assumptions and so we know it with considerable confidence¹².

The Ordovician temperature-salinity trend in Fig. 2 (which is supported by the punctal evidence discussed above) is much different from that found in modern oceans (Fig. 3). Ordovician deep waters seem to have been warmer and more saline than shallower waters, by contrast to the cold deep waters of modern oceans (that is, Ordovician oceans were salinity stratified by contrast to the thermal stratification of modern oceans). The similarity of Ordovician warm saline deep waters to modern Red Sea and Persian Gulf waters indicates that the sources of warm saline deep water were probably low-latitude marginal seas (analogous to the modern Mediterranean and Red Seas and Persian Gulf) flooded during the Ordovician high sea stand³.

Concentration of continents in the equatorial to middle latitudes probably favoured formation of evaporative marginal seas in the Ordovician⁵ (Fig. 4). Contemporaneous production of cold bottom waters in polar areas was presumably diminished by warm, equable global climate⁴ to the point that the buoyancy flux¹ of cold waters was insufficient to generate bottom waters. The cool, brackish intermediate waters (PIW in Fig. 4) represented by samples from the upper portions of the Trenton palaeoslope may reflect such a reduced (and probably warmer) polar input into a warm ocean.

High sea level, concentration of continents at low to mid-latitudes, and warm, equable global climate may thus have combined to cause ocean circulation in which warm, saline deep waters dominated the deep oceans. (A similar model has been proposed by Brass *et al.*^{1,2} to explain ocean circulation during the Cretaceous, also a time of high sea level and warm global climate.) Because warm waters contain less dissolved oxygen than cold waters, this change in circulation may also have caused decreased oxygenation of deep waters and thus had a role in the accumulation of organic-rich sediments. Decreased density contrasts between shallow and deep water masses (Fig. 3) may also have caused more sluggish circulation, so that oxygen was

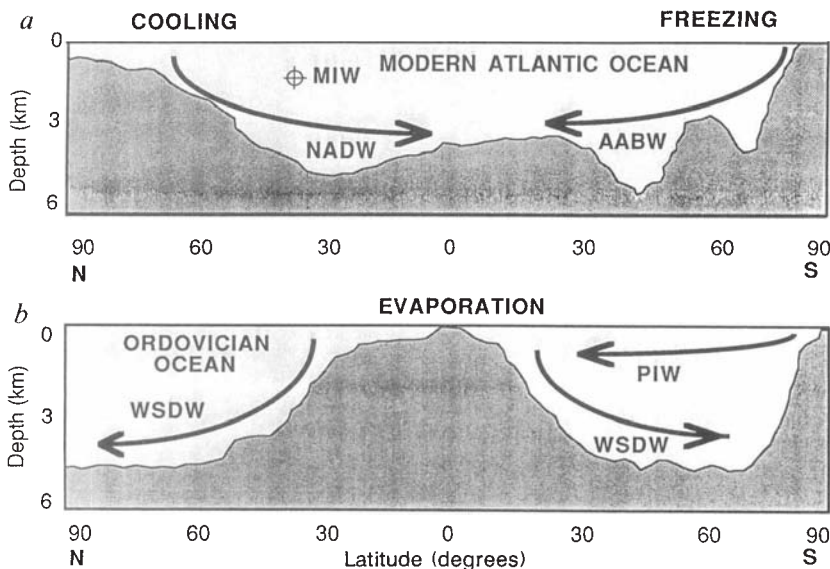


FIG. 4 *a*, Schematic north-south cross-section of modern Atlantic Ocean showing circulation of major deep water masses. Cold water masses (North Atlantic deep water and Antarctic bottom water) form at high latitudes and sink to dominate deep oceans; minor saline water mass (Mediterranean intermediate water) forms at low latitudes. *b*, Hypothesized circulation in Ordovician oceans. Warm saline deep water masses formed at low latitudes and sank to dominate deep oceans; minor cold water masses (that is, polar intermediate water) formed at high latitudes.

not replenished as readily in the deep oceans. These factors may be responsible for the abundance of Caradocian black (organic-rich) shales^{6,7}, by contrast to the paucity of similar sediments in modern, well oxygenated oceans. The circulation of cool, polar waters onto the eastern margin of the North American craton may also explain the possible 'temperate' appearance of Caradocian faunas from New York and Ontario¹⁶, despite their deposition in a tropical palaeogeographic setting⁵. □

Received August 8; accepted 17 November 1989.

1. Brass, G. W. et al. in *Climate in Earth History* (ed. Berger, W. H.) 83-89 (National Academy, Washington, DC, 1982).
2. Brass, G. W., Southern, J. R. & Peterson, W. H. *Nature* **296**, 620-623 (1982).
3. Hallam, A. A. *Rev. Earth planet. Sci.* **12**, 205-243 (1984).
4. Fischer, A. G. in *Climate in Earth History* (ed. Berger, W.) 97-104 (National Academy, Washington, DC, 1982).
5. Denham, C. R. & Scotese, C. R. *Terra Mobilis* (Earth in Motion Technologies, Houston, 1988).
6. Thickpenney, A. & Leggett, J. K. *Marine Petroleum Source Rocks, Geol. Soc. Spec. Publ. No. 26*, 231-247 (1987).
7. Hay, B. J. & Cisne, J. L. *Studies in Geology* No. 29, 113-134 (*Am. Ass. Petrol. Geol.*, 1988).
8. Rowley, D. B. & Kidd, S. F. *J. Geol.* **89**, 199-218 (1981).
9. Cisne, J. L., Karig, D. E., Rabe, B. D. & Hay, B. J. *Lethaia* **15**, 229-246 (1982).
10. Ruedemann, R. *Graptolites of North America, Geol. Soc. Am. Mem. No. 19* (1947).
11. INRS-Petrole Correlation, Source Rock Potential, and Diagenetic Evolution of the Lorraine, Utica, and Trenton Groups. *St. Lawrence Lowlands Publ. No. dp-374* (Ministère de Richesses Naturelles, Quebec, 1976).
12. Railsback, L. B. thesis, Univ. Illinois (1989).
13. Foster, M. W. *Antarctic Research Series Vol. 21* (*Am. geophys. Un.*, 1974).
14. Peck, L. S., Clarke, A. & Holmes, L. J. *Lethaia* **20**, 33-40 (1987).
15. Railsback, L. B., Anderson, T. F., Akerly, S. C. & Cisne, J. L. *Palaeoceanography* **4**, 585-591 (1989).
16. Brookfield, M. *Sedim. Geol.* **60**, 137-153 (1988).
17. Sverdrup, H. U., Johnson, M. W. & Fleming, R. H. *The Oceans: Their Physics, Chemistry, and General Biology* (Prentice-Hall, New York, 1942).
18. McGill, D. A. *Deep Sea Res.* **7**, 259-269 (1961).
19. Neumann, A. C. & McGill, D. A. *Deep Sea Res.* **8**, 223-235 (1962).
20. Seibold, E. *Geol. Rdsch.* **60**, 73-105 (1970).
21. Hay, W. W. *Geol. Soc. Am. Bull.* **100**, 1934-1956 (1988).

ACKNOWLEDGEMENTS. This research was supported by NSF grants to J.L.C. and T.F.A. and by a University of Illinois Graduate College Dissertation Research Grant to L.B.R.

Volatiles in submarine glasses as a discriminant of tectonic origin: application to the Troodos ophiolite

D. W. Muenow*†, M. O. Garcia†, K. E. Aggrey*†, U. Bednarz‡ & H. U. Schmincke‡

* Chemistry Department, University of Hawaii, Honolulu, Hawaii 96822, USA

† Hawaii Institute of Geophysics, University of Hawaii, Honolulu, Hawaii 96822, USA

‡ Institut für Mineralogie, Ruhr-Universität, Postfach 102148 Bochum, FRG

ISOTOPE ratios and concentrations of incompatible trace elements are remarkably successful in discriminating the tectonic origin and magmatic source components for basalts¹⁻⁵. But problems remain with discriminating the tectonic origin of some tholeiites, especially where field relations and other geological evidence are ambiguous. For example, the tectonic origin of basalts from the Troodos ophiolite (Cyprus) has been debated for several decades. Most workers have been unable to distinguish between an island-arc and/or back-arc origin for the ophiolite⁶⁻⁸. Here we use volatile, K_2O and TiO_2 contents from ~250 fresh submarine volcanic glasses to discriminate between tholeiites from different tectonic regimes. K_2O/H_2O ratios are lower (<0.70) in spreading-centre glasses than in those from island arcs and intraplate oceanic islands. Back-arc-basin basalts can generally be separated from mid-ocean-ridge basalts by their high H_2O contents. Using this information, we show that some fresh glasses from the Troodos ophiolite have a clear back-arc-basin affinity.

Over the past 15 years we have used high-temperature mass spectrometry for the analysis of volatiles (H_2O , Cl, F, S, CO_2 , CO, CH_4) in hundreds of specimens of submarine volcanic glasses obtained from the unaltered quenched rims of pillow and sheet-flow lavas from known tectonic environments. We have routinely used microprobe analysis of major elements to further characterize the samples. Most of these have been tholeiites; they range from primitive normal mid-ocean-ridge basalts (n-MORBs) to highly differentiated andesites and dacites from back-arc basins. We have found that the K_2O/H_2O ratio can be used to distinguish between lavas erupted at intra-plate oceanic islands (Hawaii⁹⁻¹¹) and island-arc (Mariana¹²) environments from those erupted from spreading centres (Mid-Atlantic Ridge¹³, East Pacific Rise¹⁴ and Galapagos^{15,16}) and back-arc basins (Scotia Sea¹⁷, Mariana Trough¹², Fiji/Lau¹⁸ and Woodlark basins¹⁹). Using only data for glasses that we have studied (to eliminate inter-laboratory variation), we find that distinct fields characterize each of the major tectonic environments. The potassium-rich, ocean-island-basalt (OIB) and island-arc fields ($K_2O/H_2O \geq 0.70$) contain all of the Hawaii and Mariana island-arc volcanic glasses (Fig. 1). Although H_2O abundances for these glasses range from 0.15 to 1.50 wt%, their K_2O/H_2O ratios are consistently greater than 0.70. All the data from spreading centre and back-arc-basin (BAB) glasses lie in the remaining MORB-BAB fields with $K_2O/H_2O < 0.70$. We achieve further separation of fields by plotting the K_2O/H_2O ratio against TiO_2 (Fig. 2). The titanium-rich Hawaii glasses are clearly separate from those of the Mariana island arc on this plot. We also note that in both of these plots the MORB field is contained entirely within the larger BAB field, although the MORBs have distinctly lower H_2O contents than most back-arc-basin basalts (BABs).

All of the glasses in the database for these figures are fresh (that is, alteration free) as judged by mass pyrograms and, in some cases, by stable isotopes. For example, in Fig. 3 we present two mass pyrograms (intensity as a function of temperature) that show the release pattern of water typically found for fresh and altered glasses. For the altered sample, water is released at low temperatures ($<600^\circ C$) and therefore contains a non-magmatic component to the total water abundance. Previous studies have shown that volatile-release temperatures correlate well with alteration^{10,13,20}. This has been confirmed repeatedly by carbon isotope studies which reveal an isotopically light 'contaminant' component released at temperatures below $600^\circ C$ (refs 21, 22). All of the glasses cited here release their water

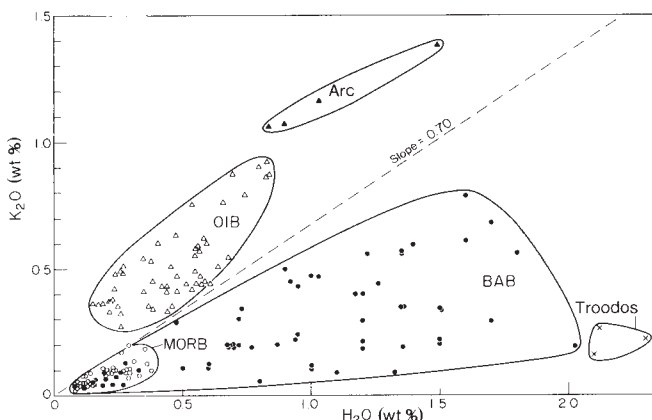


FIG. 1 Data from *Paucicrura rogata*, an orthid brachiopod in the Trenton Group of New York, along one time-parallel transect. a, Variation in density of punctae on shells with respect to reciprocal averaging score, a measure of palaeodepth⁹. b, Variation in oxygen isotopic composition of shells with palaeodepth. Letters near data clusters are used in Fig. 2 to locate samples in T-S space.

TABLE 1 Major elements and volatile abundances (wt%) in Troodos ophiolite glasses*

Sample number	316	Unaltered		Altered	
		86-9	87-46	86-6	86-34
SiO ₂	51.86	54.26	52.41	55.42	53.47
TiO ₂	0.66	0.54	0.67	1.30	1.32
Al ₂ O ₃	16.35	15.76	15.88	14.56	14.31
FeO	7.78	7.78	8.29	10.54	11.48
MgO	6.96	6.05	6.58	3.35	3.44
CaO	11.99	10.50	11.15	7.75	7.95
Na ₂ O	2.00	1.84	2.03	3.07	2.08
K ₂ O	0.16	0.22	0.26	0.29	0.68
H ₂ O	2.11	2.30	2.12	3.33	4.34
CO ₂	0.06	0.03	0.06	0.01	0.06
S	0.05	0.10	0.10	0.12	0.04
Cl	0.11	0.13	0.11	0.20	0.15
F	0.01	0.01	0.01	0.01	n.d.
Total	100.10	99.52	99.67	99.96	99.32

n.d., not detected.

* Major elements by electron microprobe (values are averages of ten spot analyses). Volatiles by high-temperature mass spectrometry. See refs 6, 27 for further sample characterization.

above 600 °C, and most of these, above 750 °C. There is also stable-isotope data for many of these glasses which indicate that these glasses are fresh. Some of the Hawaii glasses, for example, have been studied¹¹ for O and H isotopes. The δD values, $[(^2H/^1H)_{\text{sample}}/(^2H/^1H)_{\text{standard}} - 1] \times 1,000\%$, for these glasses range from -60 to -88‰ and therefore bear the typical mantle signature, despite the fact that the lavas sampled by these glasses were erupted through altered oceanic crust.

The success of K₂O/H₂O ratio as a discriminant between island-arc and back-arc-basin lavas is probably related to subduction-zone processes. The subduction of oceanic lithosphere leads to liberation of H₂O and K₂O when amphibole decomposes under the island arc. Some H₂O is probably retained within the slab as dense hydrous magnesium silicates. Under the back-arc basin this hydrated peridotite decomposes and releases H₂O to create back-arc-basin magmas²³. Some K₂O may be scavenged by the ascending fluids but the K₂O/H₂O ratio in the back-arc

remains lower than in island-arc lavas. The genetic significance for the similarity of the K₂O/H₂O ratios among island-arc and ocean-island-basaltic lavas, however, is more problematical.

We have analysed glasses from the Troodos ophiolite to determine whether volatile discrimination diagrams are useful for interpreting the origins of volcanic sequences of uncertain tectonic affinity where glass is preserved. The origin of the Troodos ophiolite has been debated^{6,24,25} for several decades. Recent studies^{6,26} have considered it to be of island-arc or back-arc origin. Geochemical characteristics of some Troodos glasses include ratios of low-field-strength elements (such as K and Rb) to high-field-strength elements (such as Ta and Ti) that are greater than those of MORBs, and Sr, Nd, Pb and O isotope signatures similar to present-day island arcs. Glasses from the chilled pillow margins of four lava flows and one dyke were analysed for volatiles using high-temperature mass spectrometry and analysed for major elements using an electron microprobe (Table 1). The glasses are from the classic north-eastern portion of the extrusive series which forms the uppermost igneous unit of the ophiolite^{6,27}. The glasses cover a wide range of compositions and are representative of all but the most mafic of the ophiolite's glass compositions^{6,7,26}. The glasses from the four lavas are strongly vesicular (up to 50 vol.%) despite having been erupted in 2 km of water²⁸.

Two of the five glasses are altered, as indicated by the large release of water beginning at low temperatures (250 °C, Fig. 3). These glasses, however, seemed fresh according to microscope examination. The remaining glasses released less H₂O (2.1–2.3 wt%) and all at higher temperatures (700–1,050 °C). They are essentially unaltered and their H₂O release patterns are identical to glass from young fresh lavas¹¹. An oxygen isotope study⁶ of some fresh Troodos glasses (including sample 316) showed that they have upper mantle $\delta^{18}O$ values (average 5.8 ‰). The one altered sample that was studied yielded a $\delta^{18}O$ value of 14‰. Previous studies^{6,7,27} of Troodos glasses have inferred high H₂O contents (2–4.5 wt%) for their magmas using the difference method (that is, 100% minus total of microprobe analysis). None of these studies was able to discriminate between low-temperature (alteration) H₂O and high-temperature (magmatic) H₂O. The high H₂O content of most of those samples (>3 wt%) probably includes non-magmatic H₂O even though the samples seemed fresh. The three unaltered glasses are H₂O-rich and Ti-poor and have low K₂O/H₂O ratios (0.08–0.12), clearly placing them within the BAB field in Fig. 2 and along an extension of the BAB field in Fig. 1. An island-arc origin for the Troodos ophiolite is also unlikely because the volcanic section has no volcanoclastic sediments, a feature characteristic of island arcs²⁸.

The high water contents of these glasses (greater than any

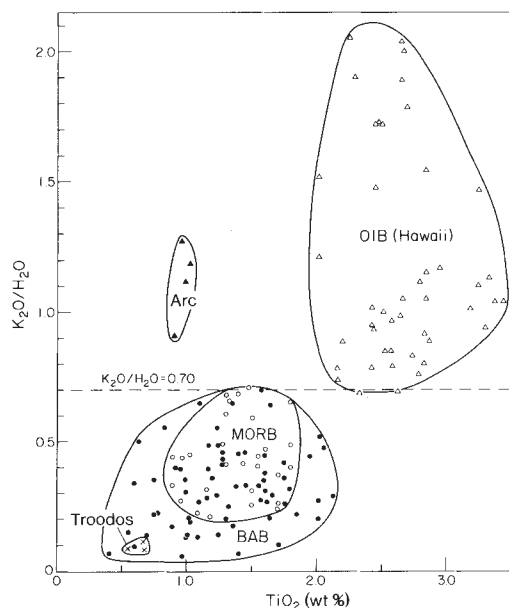


FIG. 2 K₂O/H₂O plotted against TiO₂ for unaltered submarine volcanic glasses. Data and symbols as in Fig. 1.

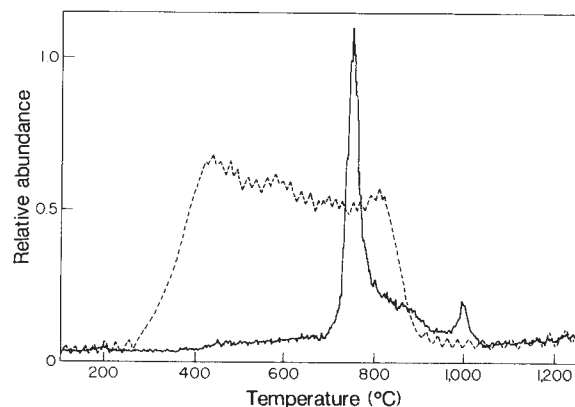


FIG. 3 Composite mass pyrogram showing H₂O release for unaltered (87-46) and altered (86-34, dashed line) Troodos glasses. (Background uncorrected.)

previously analysed basalt) combined with their low K_2O contents are diagnostic of a back-arc-basin origin. Whether these results are applicable to the entire sequence of Troodos lavas

is unclear. Nonetheless, volatile abundances seem useful as a discriminant for determining the tectonic origin of submarine lava sequences. □

Received 28 August; accepted 13 November 1989.

1. Perfit, M. R., Gust, O. A., Bence, A. E., Arculus, R. J. & Taylor, S. R. *Chem. Geol.* **30**, 227–256 (1980).
2. White, W. M. & Patchett, J. *Earth planet. Sci. Lett.* **67**, 167–185 (1984).
3. Ellam, R. M. & Hawkesworth, C. J. *Contrib. Miner. Petrol.* **98**, 72–80 (1988).
4. Pearce, J. A. & Norry, M. J. *Contrib. Miner. Petrol.* **69**, 33–47 (1979).
5. Pearce, J. A. in *Andesites* (ed. Thorpe, R. S.) 525–548 (Wiley, New York, 1982).
6. Rautenschlein, M. et al. *Earth planet. Sci. Lett.* **75**, 369–383 (1985).
7. Thy, P., Brooks, C. K. & Walsh, J. N. *Lithos* **18**, 165–178 (1985).
8. Moores, E. M., Robinson, P. T., Malpas, J., Xenophonotos, C. *Geology* **12**, 500–503 (1984).
9. Muenow, D. W., Graham, D. G., Liu, N. W. K. & Delaney, J. R. *Earth planet. Sci. Lett.* **42**, 71–76 (1979).
10. Byers, C. D., Garcia, M. O. & Muenow, D. W. *Geochim. cosmochim. Acta* **49**, 1887–1896 (1985).
11. Garcia, M. O., Muenow, D. W. & Aggrey, K. E. *J. geophys. Res.* **94**, 10525–10538 (1989).
12. Garcia, M. O., Liu, N. W. K. & Muenow, D. W. *Geochim. cosmochim. Acta* **43**, 305–312 (1979).
13. Delaney, J. R., Muenow, D. W. & Graham, D. G. *Geochim. cosmochim. Acta* **42**, 581–594 (1975).
14. Byers, C. D., Garcia, M. O. & Muenow, D. W. *Earth planet. Sci. Lett.* **79**, 9–20 (1986).
15. Byers, C. D., Muenow, D. W. & Garcia, M. O. *Geochim. cosmochim. Acta* **47**, 1551–1558 (1983).
16. Byers, C. D., Christie, D. M., Muenow, D. W. & Sinton, J. M. *Geochim. cosmochim. Acta* **48**, 2239–2245 (1984).

17. Muenow, D. W., Liu, N. W. K., Garcia, M. O. & Saunders, A. D. *Earth planet. Sci. Lett.* **47**, 272–278 (1980).
18. Aggrey, K. E., Muenow, D. W. & Sinton, J. M. *Geochim. cosmochim. Acta* **52**, 2501–2506 (1988).
19. Aggrey, K. E. thesis, Univ. Hawaii (1989).
20. Graham, D. G., Muenow, D. W. & Gibson, E. K. *Jr Proc. 10th Lunar planet. Sci. Conf.* 1617–1627 (1979).
21. Exley, R. A., Matthey, D. P., Clague, D. A. & Pillinger, C. T. *Earth planet. Sci. Lett.* **78**, 189–199 (1986).
22. Craig, H. *Earth planet. Sci. Lett.* **82**, 384–386 (1987).
23. Sinton, J. M. & Fryer, P. *J. geophys. Res.* **92**, 12782–12802 (1987).
24. Moores, E. M. & Vine, F. J. *Phil. Trans. R. Soc. A* **268**, 443–446 (1971).
25. Miyashiro, A. *Earth planet. Sci. Lett.* **19**, 218–224 (1973).
26. Flower, M. F. J. & Levine, H. M. *Contrib. Miner. petrol.* **97**, 509–524 (1987).
27. Bednarz, U. thesis, Ruhr-Universität (Bochum) (1988).
28. Bednarz, U., Sunkel, G. & Schmincke, H. U. *Geol. Surv. Can.* **85**, 183–204 (1987).

ACKNOWLEDGEMENTS. We thank M. Rautenschlein for providing glass separates. This work was supported in part by the NSF and the Hawaii Institute of Geophysics.

Occurrence of magnetic bacteria in soil

Jörg W. E. Fassbinder*, Helge Stanjek† & Hojatollah Vali*‡

* Institut für Allgemeine und Angewandte Geophysik der Ludwig-Maximilians-Universität München, Theresienstrasse 41, D-8000 München 2, FRG

† Lehrstuhl für Bodenkunde der Technischen Universität München, Hohenbachernstrasse, D-8050 Freising-Weihenstephan, FRG

‡ Present address: Department of Geological Sciences, McGill University, 3450 University Street, Montreal, Quebec, Canada, H3A 2A7

ENRICHMENT of the ferrimagnetic minerals magnetite and maghemite is frequently observed in the top layer of soil horizons^{1–5}. Although both inorganic^{6,7} and organic^{8,9} processes are known to produce magnetite, magnetite in soils has been ascribed to an inorganic origin⁶. We report here the discovery of living magnetic bacteria, similar to those found in salt- and freshwater sediments^{8,10–12}, in the A horizon of a well developed soil profile in a typical meadow environment in southern Bavaria. The bacteria were detected in fresh samples using an optical microscope equipped with a rotating magnetic field¹³ and a volumetrically calibrated depression slide, permitting accurate counts of the volume density of the organisms. We suggest that magnetic bacteria and their magnetofossils¹⁴ can contribute to the magnetic properties of soils.

Near Haindlfing in the Amper valley (Bavaria) a low moor (bog) with two histic horizons (nH1, nH2, containing >30 weight per cent (wt%) organic matter) is buried by a thin (30 cm) colluvial cover of clayey material, in which a new humose aerobic topsoil (Ah horizon) and the oxidized horizon (Go) of a gley soil developed. The lower histic horizon is separated from

the upper histic horizon by the groundwater table and corresponds to the reduced horizon (Gr) of a gley soil. Input of lithogenic magnetite from eroded material is negligible, the soil is mildly acidic and becomes more so with increasing depth (pH 5.8–6.5) (Fig. 1a).

To characterize the mineralogy and grain size of the magnetic phases^{15–19}, we studied undisturbed fresh samples of the soil using a range of rock-magnetic measurements. Magnetic susceptibility X and the saturation isothermal remanent magnetization SIRM (Fig. 1b, c) decreased continuously with increasing depth, whereas the ratio of the back-field isothermal remanent magnetization at 100 mT to the SIRM, $IRM_{100\text{ mT}}/\text{SIRM}$, (Fig. 1d) reflects the presence of trace minerals with higher coercivities (such as goethite) derived from the eroded material. The IRM curves show that >95% of the SIRM is attained at 300 mT, which is the maximum coercivity of magnetite. These data together with the Lowrie–Fuller test¹⁶ result indicate that single-domain magnetite is the dominant magnetic mineral in the Ah horizon. From the volume susceptibility and the saturation remanence of the Ah horizon we estimate the magnetite content to be ~0.035 wt% (using Fig. 4.9 in ref. 18).

We collected soil samples for optical investigations at the end of February 1989. Samples from the Ah horizon contained ~60% pore volume and had an initial water content of ~40 wt%. They were stored under deionized water and left undisturbed in the ambient geomagnetic field at room temperature (21 °C). Immediately after sampling, we observed ~100 active magnetic bacteria per ml in the Ah horizon, whereas after 14 days of storage the population density had increased to ~10⁴ ml⁻¹.

For the transmission electron microscopy (TEM) examinations (JEOL 100 CX), bacteria were magnetically separated¹⁰ from fresh samples and standard negative-staining techniques used (Fig. 2a–d). The magnetic bacteria contained magnetite crystals with octahedral or elongated shapes and morphologically resembled those found in freshwater bacteria.

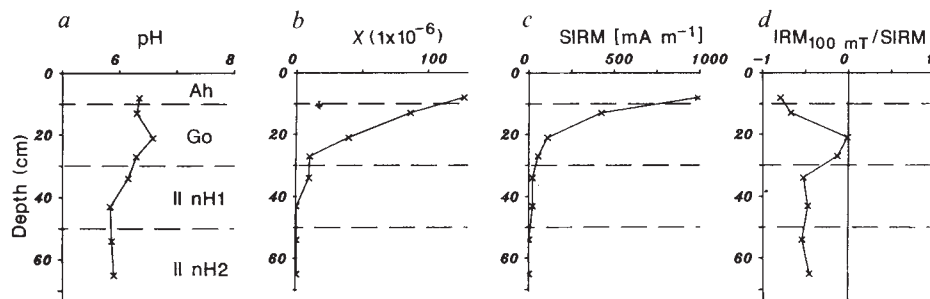
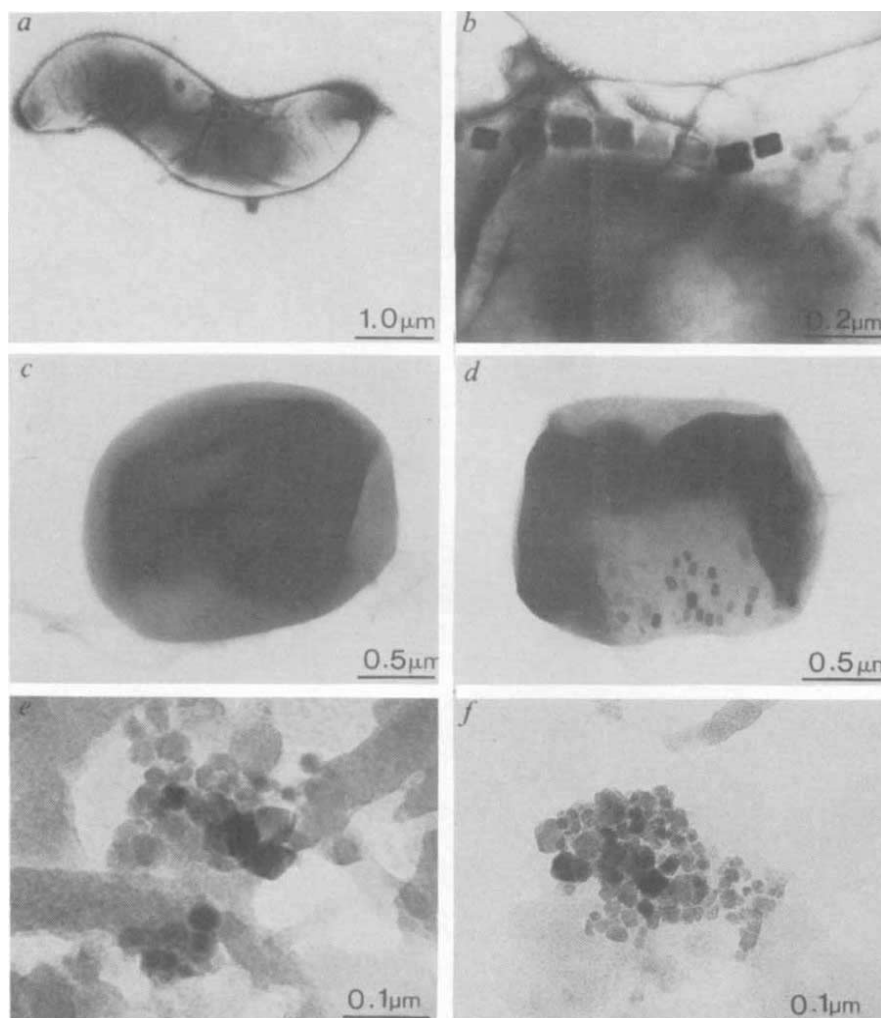


FIG. 1 a, Variation of pH (0.05M CaCl_2) with depth in the low-moor soil of Haindlfing. b, c, d, Variation of the magnetic parameters with depth in the Haindlfing low-moor soil.

FIG. 2 TEM images of uranyl-acetate-stained magnetic bacteria from the Ah horizon of the low-moor soil. *a*, A bacterium similar to *Aquaspirillum magnetotacticum*⁸ with two flagella and a single chain of ~30 magnetosomes. The magnetites have a prismatic shape. *b*, Enlargement of the magnetite chain in Fig. 2*a*. The magnetosomes have rounded edges. *c*, An unknown coccus with octahedrally shaped magnetosomes aligned in a chain. *d*, A coccus with elongated prismatic magnetosomes, which are distributed in a somewhat disorganized fashion. *e*, Magnetofossils embedded in organic material. Some of the octahedra may be assigned, on the basis of morphology, to magnetofossils whereas others display a wide range of dissolution or corrosion effects²⁴. *f*, Aggregation of magnetite crystals, which are much smaller than magnetofossils¹⁴, found in deep-sea sediments and magnetosomes in recent bacteria.



We isolated magnetic particles from the soil samples using a high-gradient extraction technique^{20,21}. TEM examination of the magnetic fraction extracted on the first day revealed mainly inactive cells of magnetic bacteria and a few particles derived from industrial fly ash (size 0.5–10 μm). The fine fraction (extracted after three days) consists of magnetites of single-domain size. The portion of the particles that was still embedded in organic material (Fig. 2*e*) or in the cell envelope, showed crystal morphologies typical of magnetofossils (size 40–100 nm). Besides these, we observed clumps of smaller magnetic particles (20–40 nm) (Fig. 2*f*). X-ray diffraction and selected-area electron diffraction (SAED) revealed the presence of magnetite (unit-cell edge length $a_0 = 8.408 \pm 0.003 \text{ \AA}$, consistent with unsubstituted pure magnetite²²) and an approximate mean crystallite dimension of 30 nm. Maghemite was not detected.

Some magnetic crystals in the fine fraction (Fig. 2*f*) were morphologically similar to those produced by the anaerobic strain of GS-15 bacteria⁹. Until now, however, extracellular magnetite derived from GS-15 bacteria has been observed only under laboratory conditions^{9,23}. Maher and Taylor have suggested an inorganic formation of similar particles in soil environments.

Because of the presence of magnetic bacteria in this soil we consider the clumps of magnetite crystals (Fig. 2*e*) to be magnetofossils showing signs of corrosion or dissolution (Fig. 2*f*). Dissolution of magnetofossils has been reported from deep-sea sediments²⁴ under oxidizing conditions as well as in freshwater sediments under reducing conditions¹⁰. Smaller magnetosomes with irregular shapes and outlines have also been observed in *Aquaspirillum magnetotacticum*^{25,26} and in undescribed natural magnetic bacteria²⁷. Magnetofossil chains similar to those

observed in deep-sea sediments²⁴ are unlikely to be preserved in soils, as they are prone to disruption by microbial activity and by freezing and drying cycles during the year. Furthermore, some magnetic bacteria do not produce magnetosomes in chains²⁷ (Fig. 2*d*).

An observed population density of only 100 magnetic cells per ml would not contribute significantly to the magnetic properties of the soil. Population density, however, is greatly influenced by environmental conditions such as water content or temperature and hence on the soil conditions at the time of sampling²⁸. A change in environmental conditions in fresh water is known to influence the production of magnetosomes and the bacterial population density^{10,27}. Under extreme conditions, where the iron and oxygen contents are unsuitable, the population of bacteria without magnetosomes rises, as was observed with pure cultures of *Aquaspirillum magnetotacticum*^{8,25}. In freshwater water sediments, population densities of magnetic bacteria of up to 10^7 ml^{-1} have been observed¹³ and in soils a total of 10^6 – 10^9 viable (nonmagnetic) bacteria per gramme have been counted²⁸. As mentioned, we observed an increase of population of the magnetic bacteria by two orders of magnitude in the laboratory in our waterlogged samples over two weeks. To extrapolate these results to *in situ* soil conditions, it would be necessary to monitor the population density of bacteria within the soil over several years.

Although we have shown that magnetic bacteria occur in our soil, it is still not clear whether or not the magnetofossils provide the dominant source of its magnetization. Nevertheless the discovery of magnetic bacteria in the A horizon of this soil may help clarify the controversial issue of the source of magnetite in soils. □

Received 1 June; accepted 20 November 1989.

1. Le Borgne, E. *Ann. Geophys.* **11**, 399–419 (1955).
2. Mullins, C. E. *J. Soil Sci.* **28**, 223–246 (1977).
3. Schwertmann, U. & Heinemann, B. *Neues Jb. Miner. Mh.* **8**, 174–181 (1959).
4. Stanjek, H. Z. *Pflanzenernähr. Bodenkunde* **150**, 314–318 (1987).
5. Fine, P., Singer, M. J., La Ven, R., Verosub, K. & Southard, R. J. *Geoderma* **44**, 287–306 (1989).
6. Maher, B. A. & Taylor, R. M. *Nature* **336**, 368–370 (1988).
7. Taylor, R. M., Maher, B. A. & Self, P. G. *Clay Miner.* **22**, 411–422 (1987).
8. Blakemore, R. P. A. *Rev. Microbiol.* **36**, 217–238 (1982).
9. Lovley, D. R., Stolz, J. F., Nord, G. L. & Phillips, E. J. P. *Nature* **330**, 252–254 (1987).
10. Vali, H., Förster, O., Amarantidis, G. & Petersen, N. *Earth planet. Sci. Lett.* **86**, 389–400 (1987).
11. Frankel, R. B., Blakemore, R. P., Torres de Aranjó, F. F., Esquivel, D. M. S. & Danon, J. *Science* **212**, 1269–1270 (1981).
12. Kirschvink, J. L. *J. exp. Biol.* **86**, 345–347 (1980).
13. Petersen, N., Weiss, D. G. & Vali, H. *Geomagnetism and Palaeomagnetism* (eds Lowes, F. J. et al.) 231–241 (Kluwer, Dordrecht, 1988).
14. Kirschvink, J. L. & Chang, S.-B. *R. Geology* **12**, 559–562 (1984).
15. Maher, B. A. *Phys. Earth planet. Inter.* **42**, 76–92 (1986).

16. Lowrie, G. W. & Fuller, M. J. *geophys. Res.* **76**, 6339–6349 (1971).
17. King, J., Banerjee, S. K., Marvin, J. & Özdemir, Ö. *Earth planet. Sci. Lett.* **59**, 404–419 (1982).
18. Thompson, R. & Oldfield, F. *Environmental Magnetism* (Allen & Unwin, London, 1986).
19. Butler, R. F. & Banerjee, S. K. *J. geophys. Res.* **80**, 4049–4058 (1975).
20. von Dobeneck, T., Petersen, N. & Vali, H. *Geowissenschaften in unserer Zeit* **5**, 27–35 (1987).
21. Petersen, N., von Dobeneck, T. & Vali, H. *Nature* **320**, 611–615 (1986).
22. Schmidt, E. R. & Vermaas, F. H. S. *Am. Mineral.* **40**, 422–431 (1955).
23. Lovley, D. R., Stolz, J. F., Maher, B. A. & Taylor, R. M. *Nature* **340**, 106 (1989).
24. Vali, H. & Kirschvink, J. L. *Nature* **339**, 203–206 (1989).
25. Gorby, Y. L., Beveridge, T. S. & Blakemore, R. P. *J. Bact.* **170**, 834–841 (1988).
26. Mann, S., Frankel, R. B. & Blakemore, R. P. *Nature* **310**, 405–407 (1984).
27. Vali, H. & Kirschvink, J. L. in *Iron Biomineralization* (eds Blakemore, R. P. & Frenkel, R. B.) (in the press).
28. Lethbridge, G. & Lynch, J. M. in *Essays in Agricultural and Food Microbiology* (Wiley, New York, 1987).

ACKNOWLEDGEMENTS. We thank J. Kirschvink, U. Schwertmann and H. Soffel for discussions and N. Petersen and H. Petermann for help with the microscope equipment. We also thank L. Bachman in whose laboratory the TEM investigations were carried out.

Earliest-known uniramous arthropod

Richard A. Robison

Department of Geology, University of Kansas, Lawrence, Kansas 66045, USA

THE earliest-known fully terrestrial animals are myriapods from the Upper Silurian of Britain¹, and unnamed myriapod-like fossils have recently been reported from Lower Silurian marine deposits in Wisconsin^{2,3}. Myriapod-like burrows in Upper Ordovician soils from Pennsylvania⁴ indicate the existence of earlier land animals. Here I report the discovery of a myriapod-like fossil from marine deposits in the Middle Cambrian Wheeler Formation of Utah that extends the record of myriapod-like body fossils back by approximately 100 Myr. It thus represents the earliest-known uniramous arthropod, and may have special significance with respect to the ancestry of the terrestrial myriapods (centipedes and millipedes) and insects.

The new myriapod-like fossil (Fig. 1a, b) is compressed in medium dark-gray, argillaceous, lime mudstone. A parting surface in the rock matrix continues horizontally through the fossilized body exposing mostly internal rather than external anatomy. Similar preservation is seen in many Cambrian fossils from the Burgess shale of the Stephen Formation in British Columbia⁵ and from the Spence, Wheeler, and Marjum formations of Utah^{6,7}. The rear half of the fossil has been mostly obliterated by weathering.

The combination of characters is unique among described arthropods and the specimen is assigned here to the new genus and new species *Cambropodus gracilis*. Like typical myriapods, it has a head with an anterior pair of long slender antenniform limbs (H1 in Fig. 1c) and an elongate trunk with several leg-bearing segments. Unlike myriapods, two pairs of postantennal head limbs (H2 and H3 in Fig. 1c) are long and slender like the trunk legs, rather than short and specialized for feeding. Proximal ends of the cephalic limbs are not exposed and details of the distal ends are not well preserved. Internally, the sclerotized head capsule shows weak vestiges of possible ligamental structures surrounding the ill-defined anterior end of a median gut trace. At least nine undifferentiated pairs of long uniramous legs are preserved along the anterior half of the trunk (T1–T9 in Fig. 1c). The curling of some limbs is similar to that seen when specimens of *Scutigera*, the common house centipede, are drowned. Based on form and sharp bending at inferred joints, the trunk legs of *C. gracilis* appear to have at least four podomeres and probably as many as six. Lengths of the second and third cephalic limbs are similar to those of the trunk legs but lengths of homologous podomeres in the head and trunk limbs may differ. Also, limbs appear to be more centrally attached beneath the head and more laterally attached on the trunk. From within the head, the gut trace continues back

through the preserved portion of the trunk where it is best seen as a broad, light-coloured, median strip along the dorsal part (Fig. 1b). Digestive diverticula are not evident and the midgut was probably very long. Minimum body length, excluding appendages, is estimated to have been ~20 mm. The complete trunk probably had at least fifteen pairs of legs.

C. gracilis is from a platy limestone and shale lithofacies representing deposition in a protected, shallow-subtidal environment near the lee side of peritidal shoals⁸. Regional distribution of lithofacies indicates that the depositional site was on the outer edge of a broad carbonate platform and was more than 250 km from the nearest coast. Fossils are generally sparse at the locality, but extensive collecting has produced a diverse biota including cyanobacteria, green algae, radiolaria, sponges, brachiopods, priapulids, pogonophorans, trilobites, phyllocarids, mollusks, echinoderms and traces of a possible cnidarian. The depositional setting and associated biota indicate that *C. gracilis* lived at or near where it was buried and in a marine environment.

Arthropod legs adapted for speed differ noticeably from those adapted for strength⁹. Those suited to fast movement are long and slender in both aquatic and terrestrial habitats^{10,11}. Those suited to slow, powerful movement are short. Therefore, *C. gracilis* was probably epibenthic and cursorial.

The method of respiration used by *C. gracilis* is not clear. In aquatic biramous arthropods such as crustaceans the outer ramus is usually a gill. Uniramous terrestrial arthropods such as insects and myriapods commonly respire by means of tracheae. Many small arthropods simply respire through a thin exoskeleton and have no special respiratory organs. The latter condition seems most probable for *C. gracilis*.

Cambropodus gen. nov.

Etymology. *Cambropodus*: an arbitrary combination of letters. The name alludes to the Cambrian age and arthropodan affinity of the animal.

Type species. *C. gracilis*.

Diagnosis. As for *C. gracilis*, see below.

Cambropodus gracilis sp. nov.

Etymology. *gracilis* (Latin): slender. In reference to the long, slender, uniramous legs of the animal.

Holotype. Part and counterpart, University of Kansas Museum of Invertebrate Palaeontology No. 204775.

Locality. NE1/4 NW1/4 sec. 20, T. 15 S., R. 10 W. (Drum Mountains Well, 7.5' quadrangle map, US Geological Survey, 1971); Drum Mountains, Millard County, Utah.

Horizon. 15.5 m below the top of the Wheeler Formation of middle Middle Cambrian age (lower *Bolaspidella* Zone^{12,13}).

Diagnosis. Exoskeleton thin and uncalcified. Head subquadrate with three pairs of long, slender, uniramous limbs; anterior pair antenniform, second and third pairs similar to trunk legs. Trunk elongate; approximately the anterior half preserving nine undifferentiated pairs of long, slender, uniramous legs. Each

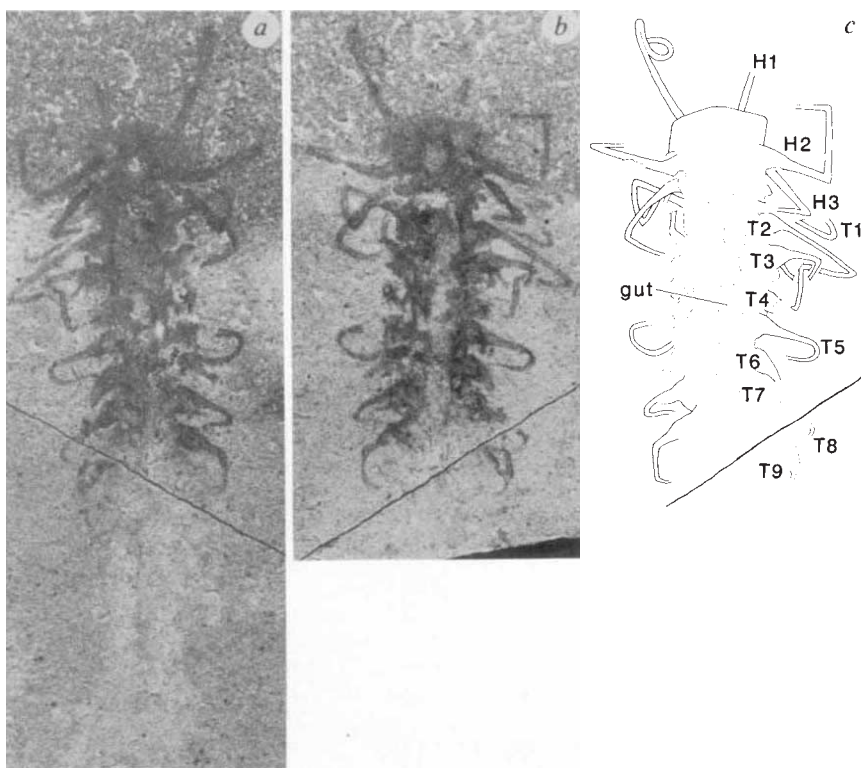


FIG. 1 *Cambropodus gracilis* gen. et sp. nov. *a*, Holotype counterpart and *b*, part, magnification $\times 7$; note that such features as the gut trace differ in detail on the part and counterpart. *c*, Explanatory drawing of part (compare with *b*); appendages are numbered from the anterior. H, head; T, trunk.

trunk leg probably containing six podomeres, with proximal podomere (possibly a coxa) distinctly swollen.

Comparative remarks. Among Cambrian fossils, the only well-documented and commonly cited uniramous animals are *Aysheaia*¹⁴ and *Xenusion*¹⁵, which are lobopodians^{15,16} rather than arthropods. *Cambropodus* differs from both *Aysheaia* and *Xenusion* in several features, including body form, presence of anterior antenniform appendages, jointed walking legs rather than lopopods, and no evidence of ringed body and limb surfaces. *Aysheaia* has been variously cited as a possible or probable marine ancestor of the terrestrial uniramous arthropods^{17,18}, but this may have been due at least in part to a lack of likely alternatives. *Cambropodus* seems to be a closer possible ancestor for these arthropods because of such shared derived characters as antenniform appendages on the head and jointed uniramous walking legs on the trunk.

In body form, *Cambropodus* superficially resembles *Portalia* and *Redoubtia* from the Burgess shale, which were originally identified as holothurians^{19,20}. Other assignments have been proposed, but the most recent evaluation²¹ concluded that their relationships are problematic. *Portalia* is based on a single incomplete specimen with several pairs of flexuous appendages that appear to bifurcate rather uniformly in the middle and lack the sharp joint-like bends present in *Cambropodus*. Moreover, appendage length and body width in *Portalia* are approximately equal, whereas leg length in *Cambropodus* is about twice the body width. It has been questioned whether the bifurcation of appendages in *Portalia* is original or the result of preservational factors such as decay²¹. From personal observations, legs of arthropods commonly separate at the podomere joints and do not split longitudinally during decay. If the splitting of appendages in *Portalia* is the result of decay, it would seem to argue against an arthropod affinity, proposed here for *Cambropodus*. The single presently accepted specimen of *Redoubtia* is poorly preserved²¹ but its appendages are much more numerous and

more closely spaced along the body than in *Cambropodus* and they show little or no proximal swelling.

Most features of *Cambropodus* are primitive for the Arthropoda. Therefore, it may be closely related to the ancestral arthropodan stock. In general body form it most closely resembles some myriapods^{9,22} but differs significantly from members of that group by its apparent lack of mandibles. Suprageneric assignment of *Cambropodus* within the Arthropoda is deferred here because of incomplete preservation and limited knowledge of the single available specimen. □

Received 2 August; accepted 3 November 1989.

1. Almond, J. E. *Phil. Trans. R. Soc. B* **309**, 227–237 (1985).
2. Mikulic, D. G., Briggs, D. E. G. & Kluessendorf, J. *Science* **228**, 715–717 (1985).
3. Mikulic, D. G., Briggs, D. E. G. & Kluessendorf, J. *Phil. Trans. R. Soc. B* **311**, 75–85 (1985).
4. Retallack, G. J. & Feakes, C. R. *Science* **235**, 61–63 (1987).
5. Whittington, H. B. *The Burgess Shale* (Yale University Press, New Haven, 1985).
6. Briggs, D. E. G. & Robison, R. A. *Paleont. Contr. Univ. Kans.* **111**, 1–23 (1984).
7. Conway Morris, S. & Robison, R. A. *Paleont. Contr. Univ. Kans.* **122**, 1–48 (1988).
8. Rees, M. N. thesis, Univ. Kansas (1984).
9. Manton, S. M. in *The Arthropoda: Habits, Functional Morphology and Evolution* (Clarendon, Oxford, 1977).
10. Clarke, K. U. *The Biology of the Arthropoda* (Elsevier, New York, 1973).
11. Hessler, R. R. in *Locomotion and Energetics in Arthropods* (eds Herreid, C. F. & Fournier, C. R.) 9–30 (Plenum, New York, 1981).
12. Robison, R. A. *J. Paleont.* **38**, 510–566 (1964).
13. Robison, R. A. *Brigham Young Univ. Geol. Stud.* **23**, 93–109 (1976).
14. Whittington, H. B. *Phil. Trans. R. Soc. B* **284**, 165–197 (1978).
15. Dzik, J. & Krumbiegel, G. *Lethaia* **22**, 169–181 (1989).
16. Robison, R. A. *J. Paleont.* **59**, 226–235 (1985).
17. Manton, S. M. in *Arthropod Phylogeny* (ed. Gupta, A. P.) 387–465 (Van Nostrand Reinhold, New York, 1979).
18. Ranzi, S. in *Myriapod Biology* (ed. Camatini, M.) 345–351 (Academic, New York, 1979).
19. Walcott, C. D. *Smithson. misc. Collns* **63**, 3–20 (1918).
20. Walcott, C. D. *Smithson. misc. Collns* **85**, 1–46 (1931).
21. Briggs, D. E. G. & Conway Morris, S. in *Problematic Fossil Taxa* (eds Hoffman, A. & Nitecki, M. H.) 167–183 (Oxford University Press, New York, 1986).
22. Hoffman, R. L. in *Treatise on Invertebrate Paleontology, Part R, Arthropoda* (ed. Moore, R. C.) 572–606 (Geological Society of America, Boulder, and University of Kansas, Lawrence, 1969).

ACKNOWLEDGEMENTS. The myriapod-like fossil was collected by V. G. Gunther. Comments were provided by J. E. Almond, L. E. Babcock, D. E. G. Briggs, S. Conway Morris and J. T. Hannibal. This research was supported by the National Science Foundation.

Does the brain know the physics of specular reflection?

Andrew Blake* & Heinrich Bülthoff†‡

* Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK

† Center for Biological Information Processing, Department of Brain and Cognitive Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

‡ Present address: Department of Cognitive and Linguistic Sciences, Brown University, Box 1978, Providence, Rhode Island 02912, USA

IMAGES of artificial and natural scenes typically contain many highlights generated by mirror-like reflection from glossy surfaces. Until recently, computational models of visual processes have tended to regard highlights as obscuring the structure of the underlying scene. The truth is that, on the contrary, highlights are rich in local geometric information. Here we report that the three-dimensional appearance of a highlight on a computer-simulated stereoscopic curved surface affects observers' judgment of surface gloss. We also show that the 3-D appearance of a highlight affects the perception of surface curvature—that is, it can force an ambiguous convex-concave figure to change state. We thus conclude that human visual analysis seems to employ a physical model of the interaction of light with curved surfaces, a model firmly based on ray optics and differential geometry.

According to ray optics, the virtual image of a light source—a specularity—is located in depth behind a glossy convex surface and, generally, in front of a concave surface (Fig. 1). But what the eye actually registers is not depth but 'relative disparity' (Fig. 2). The ray-optic 'specular stereo' model^{1,2} (Fig. 2) shows that the relative disparity of a specularity and the shape of the underlying surface are directly related. Suppose that the specular stereo model were fully utilized by the visual system: then the

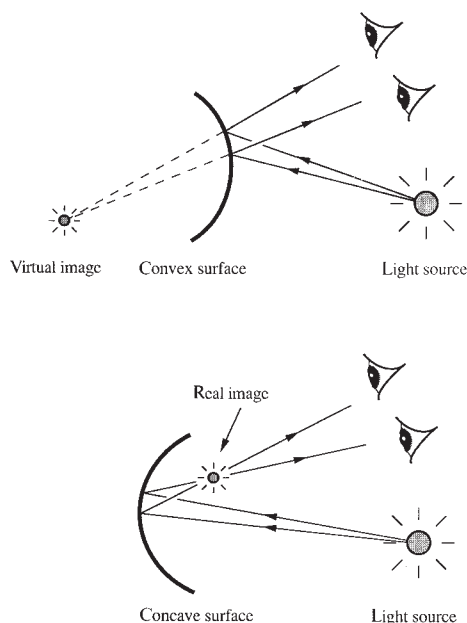


FIG. 1 The basic principle of the specular stereo model: unlike surface points such as scratches, specularities are not constrained to lie on surfaces. They appear behind a convex reflective surface but (generally) in front of a concave one. In fact things are more complicated than that: if the glossy surface is hyperbolic, specularities can appear either behind or in front of the surface^{2,10}. Generally their apparent depth varies as the surface is rotated about the line of sight. Moreover, specularities do not obey the 'epipolar' constraint¹¹ of stereoscopic vision, an important feature of the specular stereo model¹ which further complicates the idea of perceived depth¹².

relative disparity of a specularity would be consistent only with certain values of local surface curvature, if the position of the light source was known. Even without knowledge of the light-source position, the relative disparity of a specularity still constrains curvature. No convex surface can generate a convergent (−) relative disparity; a concave surface does not generate a divergent (+) one (except under certain conditions¹, unusual in practice). We performed experiments aimed to test whether the human visual system exploits such constraints.

An adjustment task was devised in which the subject interactively varied the disparity of a specularity to maximize the perceived glossiness of a stereoscopically displayed surface. Images of glossy textured curved surfaces were generated with a computer graphics workstation (Symbolics) and displayed on a high-resolution colour monitor with a stereo viewing system (Stereo-optic)³. The texture is of sufficient density to furnish strong cues for curvature from edge-based stereo. Simulation of surface gloss⁴ causes a specularity to appear superimposed on the texture, as in Fig. 3a. When the specular relative disparities are consistent with the geometry of the surface that is visible to edge-based stereo, the whole surface (not just the vicinity of the specularity⁵) appears glossy as in Fig. 3a. However, when relative disparity of a simulated specularity is inconsistent the surface ceases to look glossy (Fig. 3b).

Five test surfaces were used in the adjustment task—a convex sphere, two convex ellipsoids and two concave ellipsoids. In each case, naive subjects were asked to adjust the disparity of the simulated specularity so as to achieve the most realistic-looking glossy surface.

Results for seven subjects with the convex sphere are given in Fig. 3c. They show good agreement with the model. If the

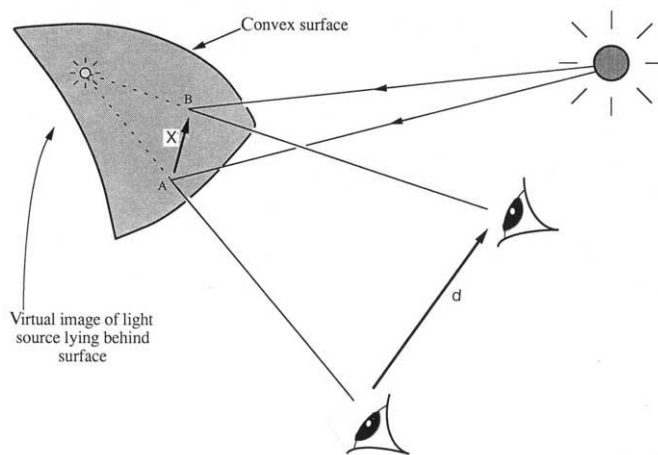


FIG. 2 Ray optics establishes a direct relationship between surface shape and the stereoscopic 'relative disparity' of a specularity^{1,2,12,13}. Disparity is the difference between angular positions of a particular feature in left and right stereoscopic views. The relative disparity of a specularity is the difference between its disparity and that of a 'surface' point lying along a common line of sight. It is a projection onto image coordinates of the displacement vector $\mathbf{x}$ between the points of reflection A and B at which light rays strike the surface. The displacement vector $\mathbf{x}$, in turn, is related linearly to baseline vector $\mathbf{d}$, the coefficients of the relation being a function solely of surface and viewer geometry². If the visual system 'knows' the physics of specular reflection and the approximate light source position, then the observed relative disparity of a specularity can be consistent only with certain values of local surface curvature. If, for example, the surface is strongly curved, then for a given baseline $\mathbf{d}$, the displacement $\mathbf{x}$ will be small, whereas for a plane surface $\mathbf{x}$ tends to be larger. The relative directions of $\mathbf{x}$ and $\mathbf{d}$ also depend on surface geometry. For a specularity on a curved surface, with given surface orientation and baseline $\mathbf{d}$, displacement $\mathbf{x}$ is in one direction if the surface is convex but in exactly the opposite direction if the surface is reversed to become concave.

visual judgements were not made according to the specular stereo model, it might be expected that subjects would attempt to place the simulated specularity on the surface. That hypothesis fails at the 99% significance level. The sign of the relative disparity of the specularity after adjustment is also correct—subjects placed the specularity behind, not in front of, the surface. Pooled results (Fig. 3c) show just a small bias towards zero disparity. It amounts to a reduction of 10–25% (99%

confidence interval) in the adjusted depth of the specularity behind the surface, compared with the relative depth predicted by the model. Results are almost as good for the other two convex ellipsoids. The sign of the relative disparity was correct for all but one out of six trials (three subjects, two different ellipsoids). Subjects did not place the specularity on the surface, and the sign of relative disparity was correct. Similar results could not be obtained for concave surfaces, however. Two

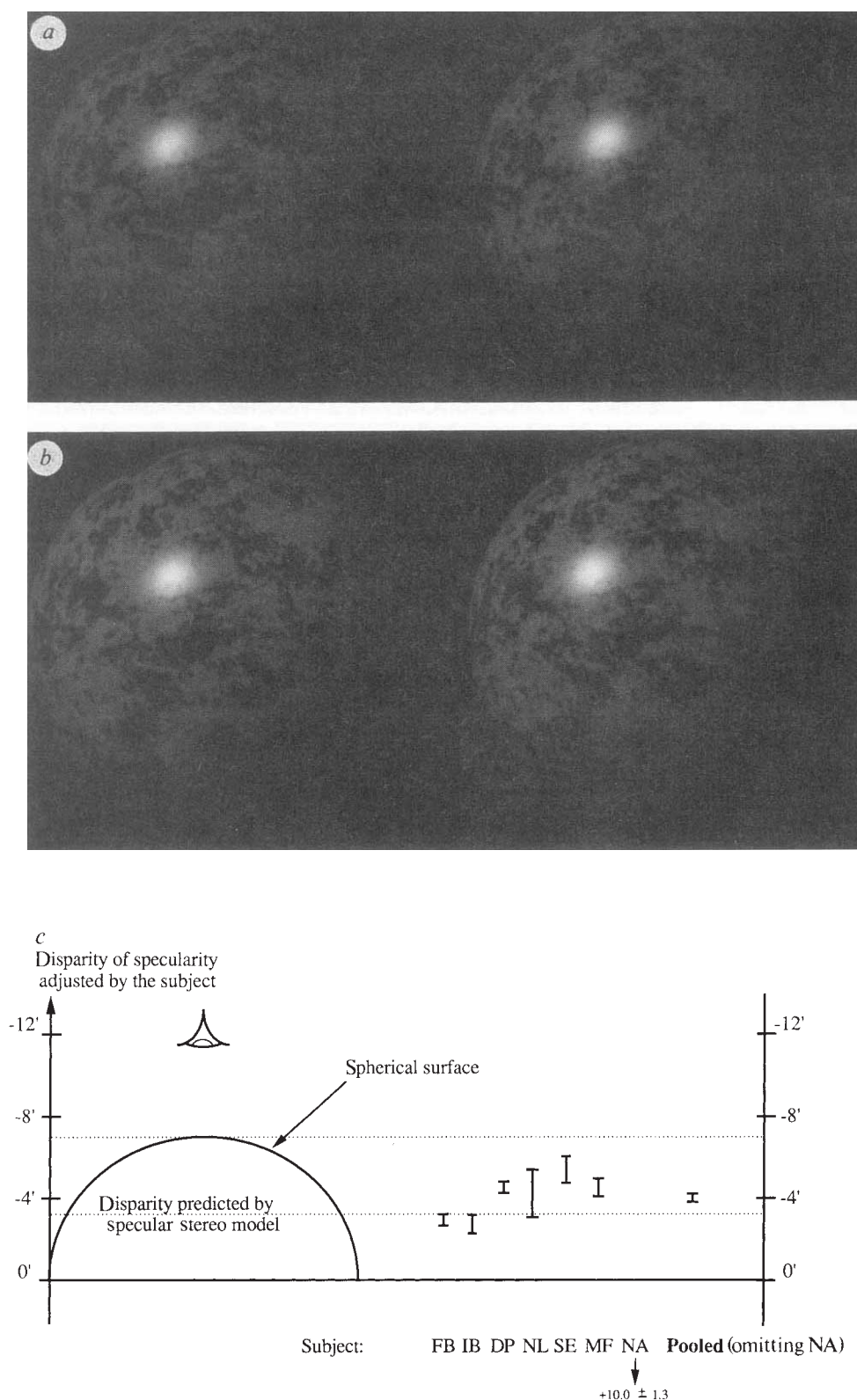


FIG. 3 The perception of surface properties can be changed by moving a simulated specularity, in depth, relative to the surface. The surface of the sphere (a) (uncrossed stereo view) looks glossy because the highlight is in the correct position behind the surface. If the highlight is in front of the surface (b)—excessively convergent (–) relative disparity—surface quality is reported to be matt and opaque, with a puff of cloud in front of it. With excessively divergent (+) relative disparity, subjects usually report that the surface looks transparent, with a light source behind it. When the relative disparity is zero the simulated specularity looks like a powdery surface patch. In an informal 2AFC experiment, 11 out of 12 naive observers who were asked which of a and b was the “polished” surface chose a. Seven naive observers performed an adjustment task in which they varied the relative disparity of a simulated highlight on a spherical surface. c, Results, giving mean and s.e.m. of responses, show a considerable degree of consistency with the specular stereo model. Four subjects (F.B., I.B., N.L., S.E.) made adjustments consistent with the model (99% significance). Two others (D.P., M.F.) showed a small bias towards zero disparity, that is, the specularity is displaced slightly towards the surface. The remaining subject (N.A.) seemed to be performing a different task, placing the specularity very far behind the surface. The pooled results show a small but significant (99%) bias of about 0.5' towards zero disparity. Pooled results show that subjects do not, on average, place the specularity on the surface.

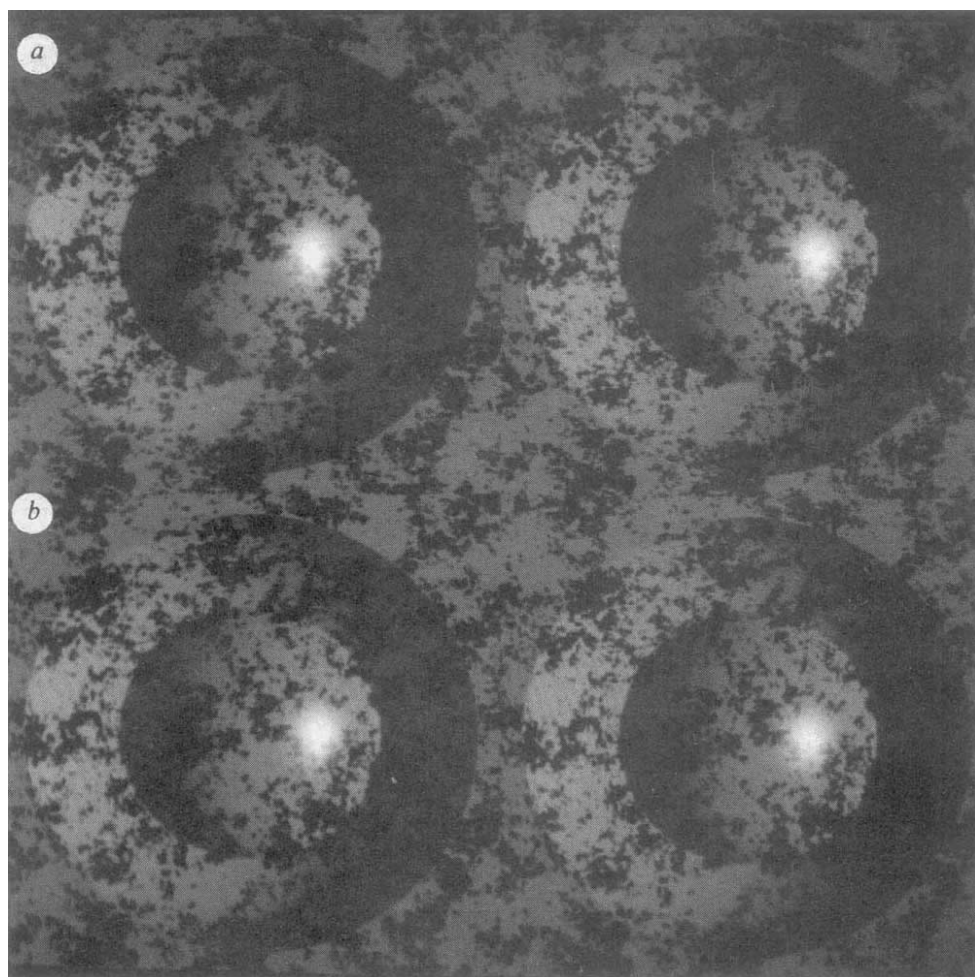
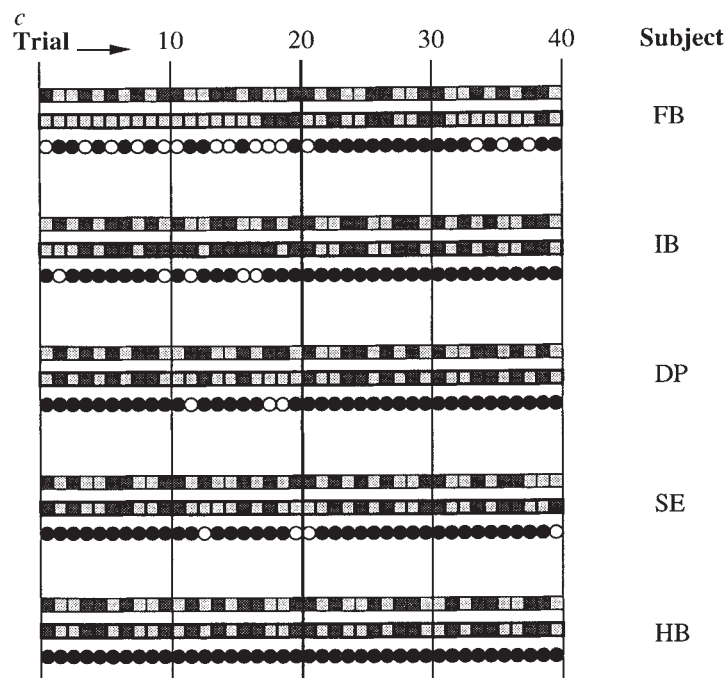


FIG. 4 The perception of surface curvature can be affected by the position of a specular highlight. As a demonstration that the human visual system 'knows' the physics of ray optics we used an image of a surface whose 3-D interpretation can flip between two states (convex, concave). If a highlight is added to the image the 3-D interpretation, in a monocular view, is biased so that the inner part of the surface tends to be convex (like a ball in a saucer). A simulated specularity superimposed on a stereoscopic view of the surface can be either behind (a) or in front of (b) the textured surface (uncrossed stereo view). The textured surface itself is at zero disparity. The two views were presented in random order, with a stereo viewing system, using 5-, 10- or 15-s exposures separated by random-dot masking frame. Subjects made a 2AFC between a convex and concave interpretation. After a short training period (20 exposures without feedback) they predominantly made choices which conformed to the predictions of the model (c). For each of five subjects (four were naive about the purpose of the experiment), the last 20 responses were biased towards those predicted by the model, at the 99% significance level. (The effect is weakened in the printed images because of the limited contrast range of the print-medium, compared with that of the CRT monitor used for the experiment; it is essential, for the strongest effect, that the highlight looks like a real reflection of a light source.



concave surfaces were used, exactly the inverses of the convex ellipsoids used previously. Out of six different naive subjects, four placed the specularity on the surface and two behind it. The model predicts, of course, that the specularity should be in front of the surface. One possible explanation for the failure of the model to predict performance for concave surfaces is as follows. Currently available graphics tools do not simulate the effects of shadowing by extended light-sources or of mutual illumination⁶, both of which are crucial for faithful rendering of concave surfaces. Neglect of these effects produces surfaces which are concave stereoscopically but, monocularly, tend to appear convex—a cue conflict³. Hence subjects reported that it was difficult to find any adjustment for which a concave surface looked convincingly glossy.

The conclusion of this experiment is that, at least for simulated convex surfaces, the human visual system models the physics of specular reflection well enough to predict relative disparity effects. The visual system 'expects'—correctly—that a specularity lies behind, not on or in front of, a convex surface. Note that results reported here are for variation of horizontal disparity. Subjects also adjust vertical disparities⁷ on convex surfaces, in accordance with the specular stereo model⁸.

Our second experiment was complementary to our first and addressed the question of whether the visual system can accommodate variations in specular relative disparities by changing its hypothesis of surface curvature, rather than its hypothesis of glossiness.

We devised the stimulus shown in Fig. 4—a stereo textured variant of an ambiguous (reversible) shaded surface⁹. The texture elements all have zero disparity, consistent with a fronto-parallel planar surface. Nonetheless, monocular shading/texture cues are not entirely overridden³, so that subjects could usually see both convex (like a ball in a saucer) and concave (like a dog bowl) interpretations. A superimposed specularity, with either convergent or divergent relative disparity, strongly influenced the interpretation. As the specular stereo model predicts, convergent relative disparity of a specularity biases the subjects' interpretation away from convex. Similarly, divergent relative disparity biases interpretations away from concave. The prediction was tested by a two-alternative forced choice (2AFC) between a convex or a concave surface interpretation, when subjects were presented with simulated specularities, divergent and convergent ($\pm 5'$), in random sequence. Note that subjects were asked whether the surface appeared convex or concave, not whether the specularity was behind or in front of the surface. Figure 4c shows how the surface interpretation predicted by the model develops gradually with repeated exposures. Although initially subjects tended to be locked into either a convex or a concave interpretation, after about 20 exposures they were usually picking, quite reliably, the interpretation that was consistent with the sign of relative disparity of the specularity. Note that the change in position of the specularity is contrary to that of the surface—when the specularity is furthest away (divergent relative disparity) the centre of the surface is nearest to the viewer (convex), and vice versa. Therefore, any explanation in terms of a pulling effect exerted by the specularity on the surface is excluded.

Thus, naive observers, asked where a specularity appears to be in relation to the surface that generated it, usually reply that it seems to lie on the surface. The first experiment that we have described attempted to show that the early visual system 'knows' better, choosing interpretations that are broadly consistent with the physics of specular reflection. The second experiment demonstrated that the stereoscopic appearance of a specularity influences perceived curvature, again in a way that is consistent with the specular stereo model. This influence was detectable despite the presence of other strong shape cues. Exactly how these cues, often in mutual conflict, are combined³ and resolved into a stable percept is unknown. □

Received 18 May; accepted 17 October 1989.

1. Blake, A. *Proc. 9th Int. Joint Conf. on Artif. Intell.* 973–976 (1985).
2. Blake, A. & Breistaff, G. J. *Proc. 2nd Int. Conf. comp. Vis.* 394–403 (Institute of Electrical and Electronics Engineers, Washington DC, 1988).
3. Bülthoff, H. H. & Mallot, H. A. *J. opt. Soc. Am.* **5**, 1749–1758 (1988).
4. Phong, B. T. *Commun. Ass. comput. Mach.* **18**, 311–317 (1975).
5. Beck, J. *Surface Color Perception* (Cornell University Press, New York, 1972).
6. Greenberg, D. P. *Science* **244**, 166–173 (1989).
7. Mayhew, J. E. W. & Longuet-Higgins, H. C. *Nature* **297**, 376–379 (1982).
8. Blake, A. & Bülthoff, H. H. *Artificial Intelligence Lab Memo* 1149 (Massachusetts Institute of Technology, Cambridge, Massachusetts, 1989).
9. Ramachandran, V. *Nature* **331**, 163–166 (1988).
10. Koenderink, J. J. & van Doorn, A. J. *Optica Acta* **27**, 981–996 (1976).
11. Mayhew, J. E. W. & Frisby, J. P. *Artif. Intell. J.* **17**, 349–385 (1981).
12. Koenderink, J. J. & van Doorn, A. J. *Biol. Cybern.* **21**, 29–35 (1976).
13. Zisserman, A., Giblin, P. & Blake, A. *Image Vis. Comput.* **7**, 38–42 (1989).

ACKNOWLEDGEMENTS. We acknowledge the support of a grant from the Office of Naval Research, Cognitive and Neural Sciences Division (H.H.B.) and the Fellowship of Engineering, Exeter College, Oxford, University of Oxford and the Science and Engineering Research Council (A.B.). We thank N. Gryzwacz, D. Kersten, D. Plummer, T. Poggio and S. Ullman for discussions.

No evidence for illegitimate young in monogamous and polygynous warblers

Ulf B. Gyllenstein^{*†}, Sven Jakobsson[‡] & Hans Temrin[‡]

^{*} Department of Biochemistry, University of California, Berkeley, California 94720, USA

[‡] Department of Zoology, Division of Ethology, University of Stockholm, S-10691 Stockholm, Sweden

IN animals with internal fertilization, paternity is uncertain. In birds, the occurrence of copulations outside the pair-bond has been documented in a number of species^{1,2}, but the extent to which these result in illegitimate young is largely unknown, and constitutes a major deficiency in our understanding of avian mating systems^{3–5}. The analysis of tandemly repeated sequences (minisatellites), has enhanced our ability to make individual identifications and paternity determinations^{6–11}. Here we describe the use of a bird minisatellite DNA probe in assigning paternity in natural populations of the monogamous willow warbler *Phylloscopus trochilus* and of the polygynous wood warbler *Phylloscopus sibilatrix*. In both species this probe detects a multiple locus pattern and a single locus that exhibits a variable number of tandem repeats¹². Although we observed intrusions by non-resident males into the territories of paired males and extra-pair copulations, no illegitimate offspring were detected among 176 young from 32 families of both species, implying that extra-pair copulations have little or no genetic impact.

Field studies were conducted in mixed deciduous and coniferous forests in central Sweden. In both species, the male changes behaviour after pair formation, shifting from singing to following the female closely around the territory (mate guarding). Females build domed nests on the ground, where a clutch of six eggs are laid, on average. The females take sole care of incubation, but both sexes feed the young during the nestling period (~13 days) as well as after fledging. The willow warbler is predominantly monogamous and polygyny is rare (<5%) in the area of our study¹³. Numerous intrusions by neighbouring or unknown males were observed into the territories of paired males. On four separate occasions, extra-pair copulations were observed, each involving cloacal contact. During the study period, a total of 26 intra-pair copulations were recorded. Polyterritorial polygyny is common for the wood warbler in central Sweden¹⁴, and on average 23% of the males become polygynously mated¹⁵. One day before the female commenced egg laying (mean, -0.92 ; s.e.m., 0.82 ; $N = 13$), her mate began singing to

[†] Present address: Department of Medical Genetics, University of Uppsala Biomedical Center, Box 589, S-75123 Uppsala, Sweden.

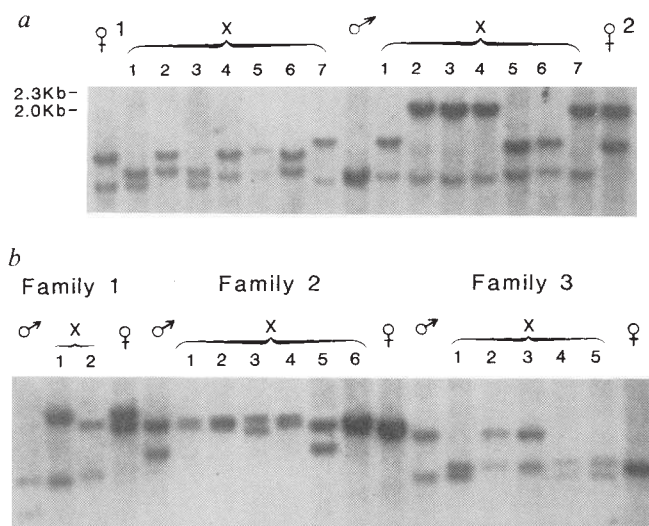


FIG. 1 *a*, Segregation of alleles at a VNTR locus in two half-sib families of willow warbler. Offspring number 7 in the mating with female number 1 has one mismatched allele. *b*, Segregation of alleles at a VNTR locus in three families of wood warbler.

METHODS. Genomic DNA (2–5 µg) was digested with *HinfI* and electrophoresed in agarose gels (0.7% for multiple locus analysis, or 1.2% for VNTR locus analysis) in 1 × TBE (0.089 M Tris-borate, 0.089 M boric acid, 0.002 M EDTA, pH 8.3). The gel was denatured and neutralized, and the DNA transferred onto a nylon membrane (Nytran)^{17,27}. Filters were hybridized with the willow warbler L17 probe¹⁷ and washed in 1 × SSPE (0.18 M NaCl, 0.01 M NaH₂PO₄, 0.001 M EDTA, pH 7.4), 0.1% SDS at 55 °C for multiple locus analysis, or in 0.1 × SSPE, 0.1% SDS at 65 °C for VNTR locus analysis.

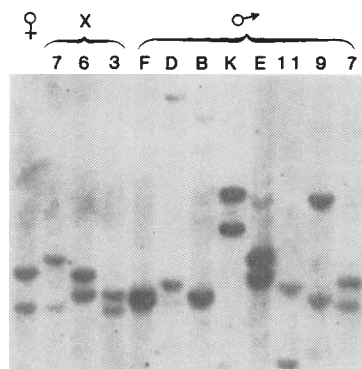


FIG. 2 Comparison of the alleles at the VNTR locus in members of family number 1 (Fig. 1*a*), including the offspring with the mismatched allele (number 7), with those of males from the surroundings (D–7). Males D, B, K and E established territories surrounding that of male F; males 11, 9 and 7 were bachelor males.

attract a second female. Eleven of the 13 males observed established a second territory 100–700 m from their first, whereas the other two sang in outlying parts of the first territory. This division of time between two territories by polyterritorial males leads to intrusions by other males, and it has been suggested that such behaviour results in a high frequency of successful extra-pair copulations in another polyterritorial species, the pied flycatcher (*Ficedula hypoleuca*)¹⁶.

Birds from 19 families (120 young) of willow warbler, 13 families (56 young) of wood warbler and 18 males from outside these families were caught in 1986 and 1987 and 5–10 µl blood was collected from them by cutting a claw. As a probe, we used a DNA sequence from the willow warbler (L17) flanking a minisatellite locus¹⁷ and which crosshybridizes to the DNA of the other bird species; this probe can detect a multiple locus pattern under low-stringency hybridization conditions and a

TABLE 1 Comparison in two species of warblers of the theoretical probabilities of genotype sharing and of false inclusion based on analysis of a single VNTR locus, a multiple locus, and a combination of both

Species	Average number of fragments scored per individual* (% fragment sharing)	Probability of genotype sharing†	Probability of false inclusion‡
Single VNTR locus			
Willow warbler§	—	1.0×10^{-3}	4.5×10^{-2}
Wood warbler	—	1.8×10^{-3}	5.9×10^{-2}
Multiple locus analysis			
Willow warbler¶	17.2 (12.1)	1.4×10^{-16}	1×10^{-7}
Wood warbler#	15.6 (12.7)	1×10^{-14}	8×10^{-7}
VNTR and multiple locus combined			
Willow warbler	—	1.4×10^{-19}	4.5×10^{-9}
Wood warbler	—	1.8×10^{-17}	4.7×10^{-8}

* The average number of minisatellite fragments scored in the range 5–24 kb in the willow warbler and the wood warbler and, in parenthesis, the fraction of minisatellite fragments shared between individuals from different areas.

† For a VNTR locus, the probability of two unrelated individuals sharing the same genotype (false association) is given by $q^2(2 - q)$ (ref. 18), where q is the average frequency of alleles. This assumes no mutation and equal allele frequencies. For both species, the observed allele frequency distribution at the VNTR locus (ref. 17, and unpublished observations), is in reasonable agreement with the assumption of equal allele frequencies¹⁸. The average allele frequency at the VNTR locus was estimated to 0.023 in willow warbler and to 0.030 in wood warbler. Because in the willow warbler, the frequency by which new length variants are generated at the VNTR locus is high¹⁷ the probability of false inclusion in Table 1 is probably an overestimate. For a multiple locus pattern, the probability of complete identity of fragment sizes in two unrelated individuals is approximated by x^y (ref. 6), where x is the proportion of shared fragments between unrelated individuals and y is the average number of fragments scored in an individual. This assumes unlinked loci (50% recombination), no mutation and an equal probability of detecting differences in migration over the whole range of fragment sizes. The probability of genotype sharing and the probability of false inclusion will increase if the recombination is less than 50% (ref. 17).

‡ For a VNTR locus, the probability of false inclusion equals $2q - q^2$ (ref. 18), assuming equal allele frequencies. For a multiple locus pattern, the probability of false inclusion for unrelated males is approximately $x^{y/2}$ (ref. 6), where $y/2$ is the minimum number of paternal-specific bands. This assumes that no maternal and paternal fragments will by chance have identical electrophoretic migration. Correct maternity is also assumed and as neither egg laying in other warbler territories nor any maternal fragment not present in the offspring has been observed, this assumption does not seem to be violated. Of the average number of fragments detected in the willow warbler (17.2), 12% or two will be shared between unrelated individuals, whereas 15.2 will be derived specifically from the two parents. On average, half of the 15.2 fragments are expected to be of paternal origin^{7,17}. If the female mates with an unrelated male, then the probability of false inclusion is therefore $\sim 0.12^{15.2/2} = 1 \times 10^{-7}$.

§ Based on analysis of the VNTR alleles of 40 willow warblers (22 parents and 18 males from various regions).

|| Based on analysis of the VNTR alleles of 19 individuals from various localities.

¶ Based on pairwise comparisons of 18 males from various localities, compared side by side in different combinations.

Based on pairwise combinations of 10 males from various localities, compared side by side in different combinations.

single locus with a variable number of tandem repeats (VNTR) at high stringency. The multiple locus fragments detectable at low stringency show mendelian inheritance and segregate as a number of unlinked or loosely linked loci; analysis is only slightly affected by new length variants¹⁷. From estimates of the average allele frequency at the VNTR locus (Fig. 1*a, b*) in warblers from outside the study area and in parental birds¹⁷, the probability of false inclusion¹⁸ is calculated to be 4.5% and

5.9% in the willow warbler and the wood warbler respectively (Table 1). The apparently high frequency by which new length variants are generated at this¹⁷ and certain other VNTR loci^{19,20} indicates that these probabilities of false inclusion may be overestimates, and that many VNTR loci may actually be suitable for paternity inclusions rather than for exclusions. The range of fragment sizes used in the multiple locus analysis (5–24 kb) does not overlap with the allelic sizes at the VNTR locus (0.5–4 kb), and the information obtained from the two analyses was therefore treated as independent, resulting in a theoretical probability of false inclusion of $\sim 10^{-8}$ for unrelated warblers (Table 1).

No illegitimate offspring were detected in any of the families by multiple locus analysis. The 95% confidence interval for the frequency of illegitimate offspring in the willow warbler is 0.0–0.03 per offspring, or 0.0–0.18 per brood. For the wood warbler, these intervals are 0.0–0.06 per offspring, or 0.0–0.25 per brood. Analysis of the VNTR locus supported our conclusion that paternity was correct in all families. No wood warbler young were mismatched relative to the alleles of their presumptive parents, although a total of seven willow warbler young were mismatches (three with a maternally derived allele, three with a paternally derived allele and one with an allele derived from either parent) (Fig. 1, offspring number 7 in the mating with female number 1¹⁷). None of the mismatched, paternally derived alleles was identified in the surrounding males (Fig. 2). Also, all the multiple locus fragments in offspring number 7 (and the other mismatched young) could be traced to the postulated male and female, indicating no illegitimacy (Fig. 3). The mismatched VNTR allele in this (and presumably all the mismatched) offspring probably represents a new length variant. The lack of new length variants in the wood warbler indicates that the VNTR locus detected in this species has a different organization or core sequence.

Correct estimates of paternity and maternity are essential for analysing the costs and benefits of different reproductive

strategies in birds. The reliability of estimates of male breeding success based on the number of fledged young in a brood has been questioned by paternity studies using polymorphisms in plumage, proteins or non-repetitive DNA, where high frequencies of extra-pair fertilizations have been found under conditions that are semi-natural^{21,22} or natural^{23–26}. We find no evidence of illegitimate young, which indicates that the number of fledged young is a reliable estimate of male breeding success in both warbler species. Similarly, paternity testing using multiple locus minisatellite analysis has been applied to a population of dunlocks (*Prunella modularis*), in which all but one of the young ($N = 133$) were fathered by resident males¹¹. Thus, DNA analyses have not substantiated high levels of extra-pair fertilization in these species. □

Received 11 September; accepted 20 November 1989.

- McKinney, F., Cheng, K. M. & Bruggers, D. J. in *Sperm Competition and the Evolution of Animal Mating Systems* (ed. Smith, R. L.) 523–545 (Academic, Orlando, 1984).
- Birkhead, T. R., Atkin, L. & Möller, A. P. *Behaviour* **101**, 101–138 (1987).
- Trivers, R. L. in *Sexual Selection and the Descent of Man* (ed. Campbell, B.) 136–179 (Aldine, Chicago, 1972).
- Mock, D. in *Perspectives in Ornithology* (eds Brush, A. H. & Clark, G. A.) 55–84 (Cambridge University Press, Cambridge and New York, 1983).
- Gowaty, P. A. *Ornithol. Monogr.* **37**, 11–21 (1985).
- Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* **314**, 67–73 (1985).
- Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* **316**, 76–79 (1985).
- Jeffreys, A. J., Brookfield, J. F. Y. & Semeonoff, R. *Nature* **317**, 818–819 (1985).
- Burke, T. & Bruford, M. W. *Nature* **327**, 149–152 (1987).
- Wetton, J. H., Carter, R. E., Parkin, D. T. & Walters, D. *Nature* **327**, 147–149 (1987).
- Burke, T., Davies, N. B., Bruford, M. W. & Hatchwell, B. J. *Nature* **338**, 249–252 (1989).
- Nakamura, Y. *et al. Science* **235**, 1616–1622 (1987).
- Jakobsson, S. thesis, Univ. Stockholm (1987).
- Temrin, H. *Anim. Behav.* **37**, 579–586 (1989).
- Temrin, H. & Jakobsson, S. *Behav. Ecol. Sociobiol.* **23**, 225–231 (1988).
- Alatalo, R. V., Gottlander, K. & Lundberg, A. *Behaviour* **101**, 139–154 (1987).
- Gyllenstein, U., Jakobsson, S., Temrin, H. & Wilson, A. C. *Nucleic Acids Res.* **17**, 2203–2214 (1989).
- Wong, Z., Wilson, V., Patel, I., Povey, S. & Jeffreys, A. J. *Ann. Hum. Genet.* **51**, 269–287 (1987).
- Wong, Z., Wilson, V., Jeffreys, A. J. & Thein, S. L. *Nucleic Acids Res.* **14**, 4605–4616 (1986).
- Jeffreys, A. J., Royle, N. J., Wilson, V. & Wong, Z. *Nature* **332**, 278–281 (1988).
- Burns, J. T., Cheng, K. M. & McKinney, F. *Auk* **97**, 875–879 (1980).
- Birkhead, T. R., Pellat, J. & Hunter, F. M. *Nature* **334**, 60–62 (1988).
- Westneat, D. F. *Anim. Behav.* **35**, 877–886 (1987).
- Sherman, P. W. & Morton, M. L. *Behav. Ecol. Sociobiol.* **22**, 413–420 (1988).
- Wrege, P. H. & Emlen, S. T. *Behav. Ecol. Sociobiol.* **20**, 153–160 (1987).
- Quinn, T. W., Quinn, J. S., Cooke, F. & White, B. N. *Nature* **326**, 392–394 (1987).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).

ACKNOWLEDGEMENTS. We thank A. C. Wilson, T. W. Quinn, C. Wiklund and O. Leimar for comments. This work was supported by the Knut and Alice Wallenberg Foundation and grants from the Swedish Natural Science Research Council to U.B.G. and the National Science Foundation to Allan C. Wilson, Department of Biochemistry, University of California at Berkeley. Henry Erlich, Department of Human Genetics, Cetus Corporation kindly provided space and support during parts of the project.

Vitronectin receptor-mediated phagocytosis of cells undergoing apoptosis

John Savill*, Ian Dransfield†, Nancy Hogg† & Christopher Haslett*

* Department of Medicine, Royal Postgraduate Medical School, Hammersmith Hospital, Du Cane Road, London W12 0HS, UK
† Macrophage Laboratory, Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, UK

PHAGOCYTE recognition of cells that have undergone apoptosis (programmed cell death) is an event of broad biological significance. Characterized by endogenous endonuclease activation¹, which results in chromatin fragmentation and nuclear condensation², apoptosis leads to swift ingestion of intact but 'senescent' or 'unwanted' cells by phagocytes in processes as diverse as the physiological involution of organs, the remodelling of embryonic tissues, and metamorphosis³. The cell-surface mechanisms by which macrophages recognize apoptotic cells as 'senescent-self' have remained obscure. Here we report that macrophage recognition of apoptotic cells (both neutrophils and lymphocytes) is medi-

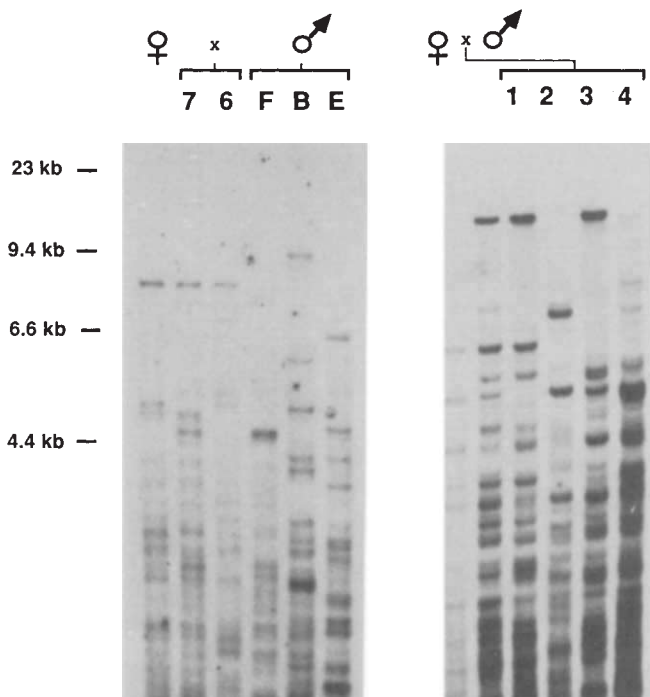


FIG. 3 Multiple locus analysis of minisatellites in two willow warbler families. Size calibration is shown on the left. Left panel: Parts of family number 1 (female 1) (as in Fig. 1a), including the offspring with the mismatched allele at the VNTR locus (7), and males from the surroundings (B, E). Right panel: Members of a family with four offspring. One new fragment length variant, not received from either parent, was found in offspring 2.

ated by the vitronectin receptor, a heterodimer belonging to the β_3 or cytoadhesin family of the integrins⁴⁻⁹. Previously, the functions of the vitronectin receptor were believed to be limited to cell anchorage¹⁰⁻¹², but our findings indicate that the receptor has a novel and direct role in self-senescent-self intercellular recognition leading to macrophage phagocytosis of cells undergoing apoptosis.

Human neutrophils 'aged' for 24 h in culture spontaneously undergo apoptosis, leading to recognition and phagocytosis of intact senescent neutrophils by human monocyte-derived macrophages^{13,14}. This interaction is specifically inhibited by

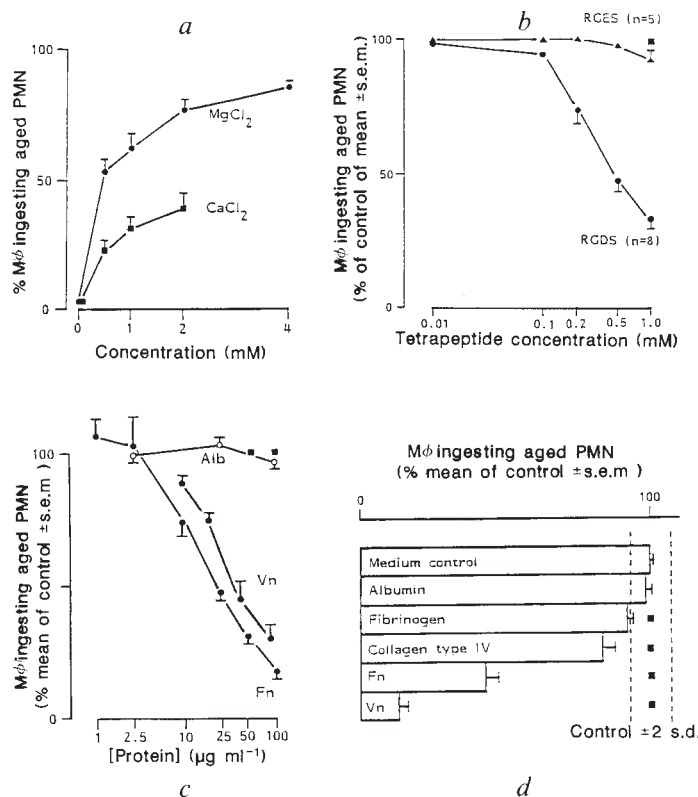


FIG. 1 Human monocyte-derived macrophage (Mφ) structures which mediate recognition of apoptotic aged neutrophils (aged PMN) exhibit integrin-like properties. *a*, Dependence on divalent cation concentration. *b*, Inhibition by Arg-Gly-Asp-Ser (RGDS), but not Arg-Gly-Glu-Ser (RGES). *c*, Inhibition by Fn and Vn in solution. *d*, Inhibition by prior attachment of macrophages to surfaces coated with Fn and Vn, but not by other RGD-bearing proteins (type IV collagen and fibrinogen). Data shown as mean ± s.e.m. (46.1 ± 2.3% of Mφ recognized aged PMN in controls).

METHODS. Human PMN were purified from blood, aged in culture for 24 h such that viability by trypan blue exclusion was >98%, and interacted with mature (5–7 day) Mφ in a 30-min serum-free phagocytic assay at 37 °C and pH 7.4 quantified by staining the washed, fixed Mφ monolayer for myeloperoxidase (absent from Mφ) to identify PMN, and then counting the proportion of Mφ ingesting aged PMN by microscopy, as previously described^{13,15}. aged PMN (5×10^6) were added in 1 ml Hank's buffered salt solution (HBSS; Gibco) to each well of Mφ cultured in 24-well tissue culture-treated plates (Falcon Plastics, USA). For *a* only, the interaction medium was HBSS without Mg^{2+} and Ca^{2+} ; >2 mM $CaCl_2$ was toxic to aged PMN (that is <98% excluded trypan blue dye). *b*, Interaction medium contained RGDS or RGES (Peninsula Laboratories, St Helens); RGDS (1 mM) had no effect on Mφ ingestion of a control particle, IgG-opsonized ox erythrocytes (ElgG; ■). *c*, Fn (from Drs M. Smith and D. Rees, and from Calbiochem, Cambridge Bioscience, UK), Vn (Calbiochem) and albumin (Alb; Sigma) were included in the interaction medium ($n=12$); Fn and Vn did not affect ElgG ingestion (■). *d*, Tissue culture wells were coated with Alb, Fn, Vn, fibrinogen (Sigma) or type IV collagen by incubation for 2 h at 21 °C with HBSS containing proteins at $80 \mu g ml^{-1}$, washed, and then Mφ (removed from adherent culture by incubation in 5 mM EDTA in HBSS for 20 min at 4 °C) were allowed to adhere to these wells in HBSS for 30 min at 37 °C to modulate surface receptors¹⁸ before interaction with aged PMN for 30 min as above ($n=12$); the 'medium control' wells were incubated for 2 h with HBSS alone. No effect on ElgG (■) ingestion was observed.

amino sugars and basic amino acids¹⁵, which can also inhibit ligand binding by members of the integrin superfamily of cell-surface heterodimers (for example, fibrinogen binding to the platelet integrin GPIIb-IIIa; ref. 16). Integrins are further implicated in the recognition mechanism because phagocytosis of aged neutrophils by monocyte-derived macrophages was dependent on Mg^{2+} and Ca^{2+} (refs 5 and 6; Fig. 1a) and specifically inhibited by the tetrapeptide Arg-Gly-Asp-Ser (RGDS; single-letter code), but not by Arg-Gly-Glu-Ser (RGES), at concentrations known to inhibit integrin binding to ligands bearing the RGD adhesion signal^{17,5,6} (Fig. 1b). Furthermore, aged neutrophil phagocytosis was specifically inhibited by the RGD-bearing proteins vitronectin (Vn) and fibrinogen (Fn), whether in solution in the interaction (Fig. 1c) or if used as a substrate on which monocyte-derived macrophages are allowed to modulate receptors¹⁸ before the interaction (Fig. 1d). Further evidence (data not shown) that the inhibitory effect of Fn and Vn is localized to the surface of the monocyte-derived macrophages and not the aged neutrophils is (1) the failure of apoptotic neutrophils to adhere to surfaces or beads coated with Fn, and (2) the lack of effect of preincubation of aged neutrophils with soluble Fn on recognition by macrophages, whereas preincubation of macrophages with Fn inhibited recognition of aged neutrophils. Thus it seems that a monocyte-derived macrophage integrin participates directly in RGD-dependent and divalent cation-dependent recognition of aged neutrophils.

Integrins of the β_3 family (that is GPIIb-IIIa and the vitronectin receptor (VnR)) have been implicated in the binding of

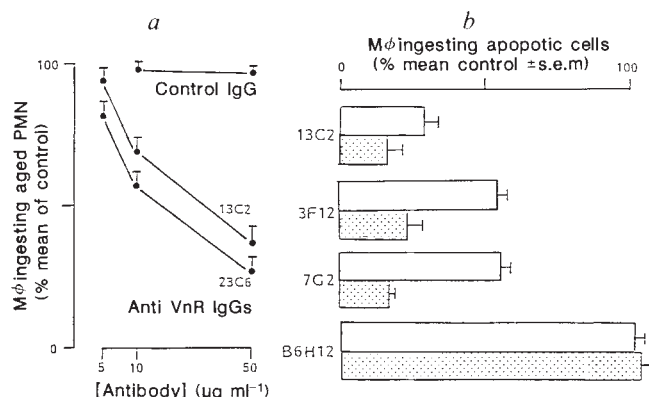


FIG. 2 Inhibition of Mφ recognition of apoptotic neutrophils and lymphocytes by monoclonal antibodies specific for the α_v - and β_3 -chains of the VnR. *a*, Specific inhibition of Mφ recognition of aged neutrophils by two monoclonal antibodies to the α_v -subunit (23C6 and 13C2)²¹; the Mφ-binding²⁹ control monoclonal antibody Mac 2-158 had no effect, indicating that inhibition was not due to nonspecific 'steric' effects ($n=12$). *b*, Inhibition of Mφ recognition of apoptotic neutrophils (open bars) and apoptotic lymphocytes (shaded bars) by monoclonal antibodies to both the α_v -subunit (13C2 and 3F12) and the β_3 -subunit (7G2)²², but not by a monoclonal antibody to the α -chain of the VnR-like LRI²² (B6H12); $n=12$. No antibody to α_v - or β_3 -subunits tested affected Mφ ingestion of ElgG (data not shown).

METHODS. Wells of washed Mφ were incubated for 15 min at 4 °C with antibody in 300 μl Iscove's medium (Gibco), warmed to 37 °C, and 100 μl of Iscove's medium containing 2.5×10^6 aged PMN added; after 15 min at 37 °C and pH 7.4, the interaction was quantified as in Fig. 1. *b*, Experiments were also performed according to the same protocol with apoptotic lymphocytes prepared by γ -irradiation (750 rad over 2 min) and overnight culture at 37 °C in RPMI 1640 with 5% FCS (Gibco) of human peripheral blood lymphocytes purified by elutriation; interaction was quantified (after washing with cold saline containing 5 mM EDTA and fixing in 2% glutaraldehyde) by staining with hemalum and counting 10 fields per well under high power ($\times 400$); 63.7 ± 3.1% of Mφ interacted with apoptotic lymphocytes; <2% interacted with non-irradiated, non-apoptotic lymphocytes cultured for the same period. Apart from 13C2 which was used at $50 \mu g ml^{-1}$, monoclonal antibodies were used at a 1-in-25 dilution of hybridoma ascites; isotype control monoclonal antibodies in the same 'vehicle' (that is, ascites or purified antibody) had no effect (data not shown). Dr M. Horton gave monoclonal antibodies 23C6 and 13C2; Dr E. Brown gave monoclonal antibodies 7G2, 3F12 and B6H12.

monocyte-derived macrophages to Vn and Fn (refs 9 and 20). Monoclonal antibodies 13C2, 23C6 and 3F12, all specific for the α_v -subunit (the α -chain of the VnR)^{21,22}, and monoclonal antibody 7G2, specific for the β_3 -subunit (the β -chain of the VnR)²², specifically inhibited phagocytosis of aged neutrophils by monocyte-derived macrophages (Fig. 2). Evidence that VnR on monocyte-derived macrophages and not on aged neutrophils mediates recognition is that (1) none of the inhibitory monoclonal antibodies specific for the α_v - or β_3 -subunits bound to aged neutrophils as assessed by flow cytometry (data not shown), and (2) neither α_v - nor β_3 -specific monoclonal antibodies (including two used in the study described here, 3F12 and 7G2) immunoprecipitate any peptides from surface-labelled neutrophils²². Furthermore, a heterodimer with electrophoretic characteristics typical of the VnR (relative molecular mass (M_r) 155,000/95,000 (155K/95K), changing to 130K/100K under reducing conditions; Fig. 3a) was immunoprecipitable from ¹²⁵I-surface-labelled macrophages by monoclonal antibodies specific for both the α_v -subunit (23C6 and 13C2) and the β_3 -subunit (7G2). Indeed, polypeptides immunoprecipitated from fibroblasts, the cell-type in which the VnR was originally described⁴, exhibited very similar electrophoretic mobilities to those immunoprecipitated from macrophages by monoclonal antibodies to both the α_v -subunit (13C2) and the β_3 -subunit (7G2) (Fig. 3b). Although the α_v -subunit may be complexed with β -chains other than the β_3 -subunit (for example, the β_x -subunit²³ and the β_1 -subunit²⁴), these are immunologically dis-

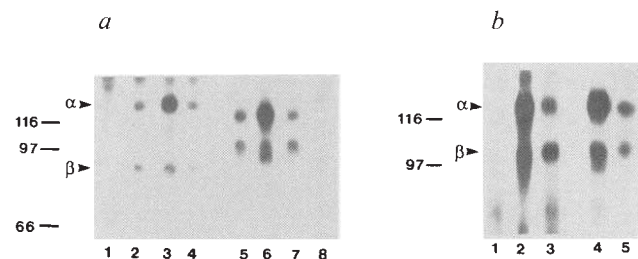


FIG. 3 Immunoprecipitation analysis of integrins of the β_3 -subunit family present on ¹²⁵I-surface-labelled macrophages and fibroblasts. *a*, Immunoprecipitates obtained from M ϕ with monoclonal antibodies against the α_v -subunit (23C6, lanes 2 and 5; 13C2, lanes 3 and 6) and the β_3 -subunit (7G2, lanes 4 and 7), and an irrelevant isotype control monoclonal antibody (lanes 1 and 8); analysis by SDS-PAGE under non-reducing (lanes 1-4) and reducing (lanes 5-8) conditions. Note similar electrophoretic mobilities of polypeptides precipitated by monoclonal antibodies to both chains of the VnR. The identity of the subsidiary band at M_r 120K in lane 3 (13C2), which possibly represents an α -chain precursor, although the M_r differs slightly from that described in melanoma cells¹², is now being investigated (M. Horton, personal communication). *b*, Immunoprecipitates from the cultured human foreskin fibroblast cell line Butler (lanes 1, 2 and 3) and M ϕ (lanes 4 and 5) obtained with monoclonal antibodies against the α_v -subunit 13C2, lanes 2 and 4), and the β_3 -subunit 7G2, lanes 3 and 5), and an irrelevant isotype control monoclonal antibody (lane 1); analysis simultaneously under reducing conditions. Note similar electrophoretic mobilities of polypeptides precipitated both cell types. M_r markers, $M_r \times 10^{-3}$.

METHODS. Fibroblasts and M ϕ resuspended at 2×10^7 per ml in PBS containing 50 mM glucose were ¹²⁵I-labelled using the lactoperoxidase-glucose oxidase method³⁰ by incubation with lactoperoxidase (4.5 U ml⁻¹), glucose oxidase (0.25 U ml⁻¹) (Sigma) and 1 mCi ml⁻¹ ¹²⁵I (Amersham) for 15 min at 21 °C. Subsequent steps were performed at 4 °C. Cells were washed four times in RPMI containing 0.1% azide, and then lysed in 50 mM Tris Buffer, pH 8.0, containing 1% Nonidet P-40 (NP-40; BDH), 0.4 U ml⁻¹ Aprotinin A (Sigma), 100 mM iodoacetamide, 1 mM phenylmethylsulphonyl (PMSF), 0.15 M NaCl, and 1 mM EDTA. Lysate (1 ml) was precleared with 5 μ l normal mouse serum and 50 μ l Sepharose protein A (10% suspension; Pharmacia). Precleared lysates were incubated with the equivalent of 5 μ g monoclonal antibody for 2 h then precipitated with 50 μ l Sepharose protein A (10% suspension). Precipitates were washed twice in lysis buffer containing 0.5 mM NaCl then twice in lysis buffer and finally in 50 mM Tris buffer, pH 8.0, containing 0.05% SDS. Samples were boiled in 0.125 M Tris buffer, pH 6.8, containing 1% SDS in the absence or presence of 1% β -mercaptoethanol, and separated on 9% resolving gels.

tinct from the β_3 -subunit, so the immunoprecipitates and inhibitory effects obtained with 7G2 (anti- β_3) is strong evidence against a role for such structures in the recognition of aged neutrophils. Furthermore, none of a panel of 14 monoclonal antibodies specific to GPIIb-IIIa, which has the β_3 -subunit in common with VnR, inhibited the interaction (data not shown); neither did B6H12, a monoclonal antibody (Fig. 2) that inhibits the function of the 'leucocyte response integrin' (LRI) recently characterized on myeloid cells as having some immunological cross-reactivity with the VnR, but which is believed to have different α - and β -chains^{20,22}. These data show that macrophage VnR plays an important part in phagocytosis of apoptotic neutrophils. Indeed, because antibodies specific for α_v - and β_3 -subunits specifically inhibited recognition of apoptotic lymphocytes by monocyte-derived macrophages (Fig. 2), it seems likely that macrophage VnR has a general role in the recognition of cells undergoing apoptosis.

Indeed, there was evidence that maturation-related expression

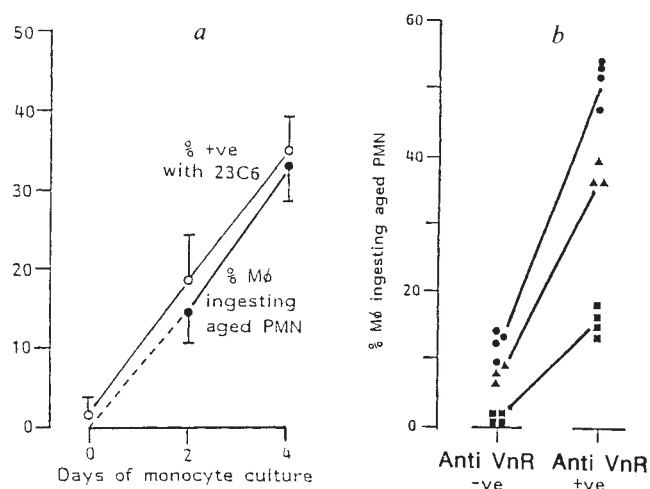


FIG. 4 Maturation-related expression of the α_v -epitope recognized by monoclonal antibody 23C6 on monocyte-derived macrophages correlates with ability to recognize aged neutrophils. *a*, Percentage of monocyte-macrophages expressing the epitope for 23C6 at 0, 2 and 4 days in culture (○) and recognizing aged neutrophils (●); the dotted line represents the fact the freshly isolated monocytes do not recognize aged PMN¹⁴. *b*, Recognition of aged PMN by 4-day-matured M ϕ separated by FACS according to expression of the epitope bound by 23C6. Three experiments (each with three to four replicates) are shown. In all experiments, M ϕ in either population were >95% viable by trypan blue dye exclusion, and >90% ingested ElgG.

METHODS. M ϕ recognition of aged PMN was assayed in the standard 30-min assay (Fig. 1). *a*, Monocyte-M ϕ were detached from adherent culture by incubation in 5 mM EDTA in HBSS for 30 min at 4 °C, washed, and then 5×10^5 cells incubated with monoclonal antibodies 23C6, P3, or Mac 2-158 (P3 is an isotype-matched negative control, and Mac 2-158 is an isotype-matched positive control) at 10μ g ml⁻¹ in HBSS-0.1% BSA for 15 min at 4 °C, washed three times and then incubated with FITC-conjugated F(ab')₂ sheep anti-mouse immunoglobulin antibody (Sigma) at 1-in-30 dilution in HBSS-0.1% BSA for 15 min at 4 °C before three washes and fixation in 2% paraformaldehyde. Fluorescence was determined by flow cytometry (EPICS II, Coulter, Hialeah, USA), 'positive' cells scored as those with fluorescence >99% of that recorded using monoclonal antibody P3. *b*, Four-day-matured M ϕ were immunolabelled by the same methods as for *a* and separated by FACS (EPICS CS, Coulter), collecting the top 30% and bottom 30% of the bimodal fluorescence intensity distribution with 23C6. After washing, M ϕ were adhered to culture plates (5×10^4 M ϕ to each well of a 96-well plate) for 30 min at 21 °C before interaction with aged PMN (0.25×10^6 per well) or ElgG (0.5×10^6 per well) for 30 min at 37 °C. In one experiment (■) 23C6 was used at 10μ g ml⁻¹; 32.3% of cells were positive; note minimal recognition of 'negative' cells, but recognition by only 15% of positive M ϕ . That this low capacity of 23C6-positive M ϕ to recognize aged PMN was likely to reflect inhibition by the antibody was suggested by the much greater recognition in two experiments using 23C6 at 2.5μ g ml⁻¹ to determine the positive population (●, ▲).

of the VnR determines the capacity of monocyte-derived macrophages to recognize aged neutrophils. First, the proportion of monocyte/macrophages that bound 23C6 (anti- α_v) increased with time that the monocytes had been matured in culture, and showed a close relationship to the proportion of cells recognizing aged neutrophils (Fig. 4a). Second, flow cytometric fractionation of 4-day-matured monocyte-derived macrophages according to binding of 23C6 showed that macrophages expressing the α_v -subunit recognized aged neutrophils to a much greater degree than did those not expressing the α_v -subunit (Fig. 4b). This is important evidence against a major role in recognition of apoptotic neutrophils for any of the integrins borne by freshly isolated monocytes (that is, the $\alpha_5\beta_1$ fibronectin receptor (FnR)²⁵, the β_2 integrins^{26,27}, and the LRI^{20,22}). There is additional evidence against a role for macrophage β_2 integrins in aged neutrophil uptake; blockade with monoclonal antibodies to macrophage CR3 and CR4/p150,95 failed to affect macrophage phagocytosis of aged neutrophils¹⁵ and macrophages from a patient with congenital β_2 -subunit deficiency (β_2 -subunit expression, 3.8% of control) recognized aged neutrophils to a degree similar to controls²⁸. Therefore, it seems likely that the inhibitory effect of both Vn and Fn was on the macrophage VnR, rather than on other integrins borne by these cells, particularly as the FnR was not inhibited by soluble fibronectin⁶ (which did inhibit aged neutrophil recognition (Fig. 1c)). The data do not exclude an accessory role for other macrophage receptors, but they indicate that the VnR plays a major role in recognition of apoptotic neutrophils; the most potent anti-VnR monoclonal antibody (23C6) inhibited the interaction by over 70%.

In conclusion, the study presented indicates a novel and direct role for the VnR in macrophage recognition of cells undergoing apoptosis. Further clarification of the role of the VnR in recognition of senescent-self, including identification of the ligand responsible for promoting phagocytosis of apoptotic cells through the VnR, will provide additional new insights into the signalling repertoire of the RGD sequence and the integrins. Indeed, these observations will also have important implications for the understanding of the regulation of tissue injury in inflammation, because there is evidence that macrophage recognition and clearance of neutrophils undergoing apoptosis occurs at inflamed sites *in vivo*. This could represent a mechanism that limits tissue injury and promotes resolution through the removal of intact senescent neutrophils which have not disgorged their potentially harmful contents¹³. □

Received 15 August; accepted 8 November 1989.

- Wyllie, A. H. *Nature* **284**, 555–556 (1980).
- Wyllie, A. H., Morris, R. G., Smith, A. L. & Dunlop, D. J. *Pathol.* **142**, 62–77 (1984).
- Wyllie, A. H., Kerr, J. F. R. & Currie, A. R. *Int. Rev. Cytol.* **68**, 251–306 (1980).
- Pytela, R., Piersbacher, M. D. & Ruoslahti, E. *Proc. natn. Acad. Sci. U.S.A.* **82**, 5766–5770 (1985).
- Hynes, R. O. *Cell* **48**, 549–554 (1987).
- Ruoslahti, E. R. & Piersbacher, M. D. *Science* **238**, 491–497 (1987).
- Pytela, R., Piersbacher, M. D., Ginsberg, M. H., Plow, E. F. & Ruoslahti, E. *Science* **232**, 1559–1561 (1986).
- Charo, I. F., Bekeart, L. S. & Steiner, B. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 8351–8355 (1986).
- Phillips, D. R., Charo, I. F., Parise, L. V. & Fitzgerald, L. A. *Blood* **71**, 831–843 (1988).
- Charo, I. F., Bekeart, L. S. & Phillips, D. R. *J. biol. Chem.* **262**, 9935–9938 (1987).
- Cheresh, D. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6471–6475 (1987).
- Cheresh, D. & Harper, J. R. *J. biol. Chem.* **262**, 1434–1437 (1987).
- Savill, J. S. *et al. J. clin. Invest.* **83**, 865–875 (1989).
- Newman, S. I., Henson, J. E. & Henson, P. M. *J. exp. Med.* **156**, 430–442 (1982).
- Savill, J. S., Henson, P. M. & Haslett, C. *J. clin. Invest.* **84**, 1518–1527 (1989).
- Nachman, R. L. & Leung, L. L. K. *J. clin. Invest.* **69**, 263–269 (1982).
- Piersbacher, M. D. & Ruoslahti, E. R. *Nature* **309**, 30–33 (1984).
- Wright, S. D. *Meth. Enzym.* **132**, 204–221 (1986).
- Burns, G. F., Cosgrove, L., Triglia, T. *et al. Cell* **45**, 269–280 (1986).
- Brown, E. J., & Goodwin, J. L. *J. exp. Med.* **167**, 777–793 (1988).
- Davies, J. *et al. J. Cell Biol.* **109**, 1817–1826 (1989).
- Gresham, H. D., Goodwin, J. L., Allen, P. M., Anderson, D. C. & Brown, E. J. *J. Cell Biol.* **108**, 1935–1943 (1989).
- Cheresh, D., Smith, J. W., Cooper, H. M. & Quaranta, V. *Cell* **57**, 59–68 (1989).
- Frederick, G., Gallit, J., van der Geer, P., Ruoslahti, E. & Hunter, T. *EMBO J.* **8**, 2955–2965 (1989).
- Brown, D. L., Phillips, D. R., Damsky, C. H. & Charo, I. F. *J. clin. Invest.* **84**, 366–370 (1989).
- Myons, B. L., Dalzell, J. G., Hogg, N. & Ross, G. D. *J. clin. Invest.* **82**, 640–651 (1988).
- Wright, S. D. & Detmers, P. A. *J. Cell Sci. (Suppl.)* **9**, 251–306 (1988).
- Savill, J. S. thesis, Univ. London (1989).
- Morganelli, P. M. & Guyre, P. M. *J. Immun.* **140**, 2296–2304 (1988).
- Hubbard, A. L. & Cohn, Z. A. *J. Cell Biol.* **64**, 438–460 (1975).

ACKNOWLEDGEMENTS. We thank Dr M. Horton (St Bartholomew's Hospital), Dr E. Brown (Washington School of Medicine, St Louis, USA), and Dr P. Morganelli (Dartmouth Medical School, Hanover, USA) for gifts of antibodies, Dr A. Wyllie (University of Edinburgh) for advice on the preparation of apoptotic lymphocytes, Drs M. Smith and D. Rees (National Institute for Medical Research) for advice and gifts of Fn, and Dr A. J. Rees (Renal Unit, Hammersmith Hospital) for type IV collagen. J.S.S. is a MRC Training Fellow, and C.H. is a MRC Senior Clinical Fellow: support from the MRC and the Special Trustees of Hammersmith Hospital is acknowledged. I.D. and N.H. are supported by the ICRF.

Specification of limb development in the *Drosophila* embryo by positional cues from segmentation genes

Stephen M. Cohen*

Lehrstuhl für Genetik, Universität München, Maria Ward Strasse 1a, D8000 München, 19 FRG

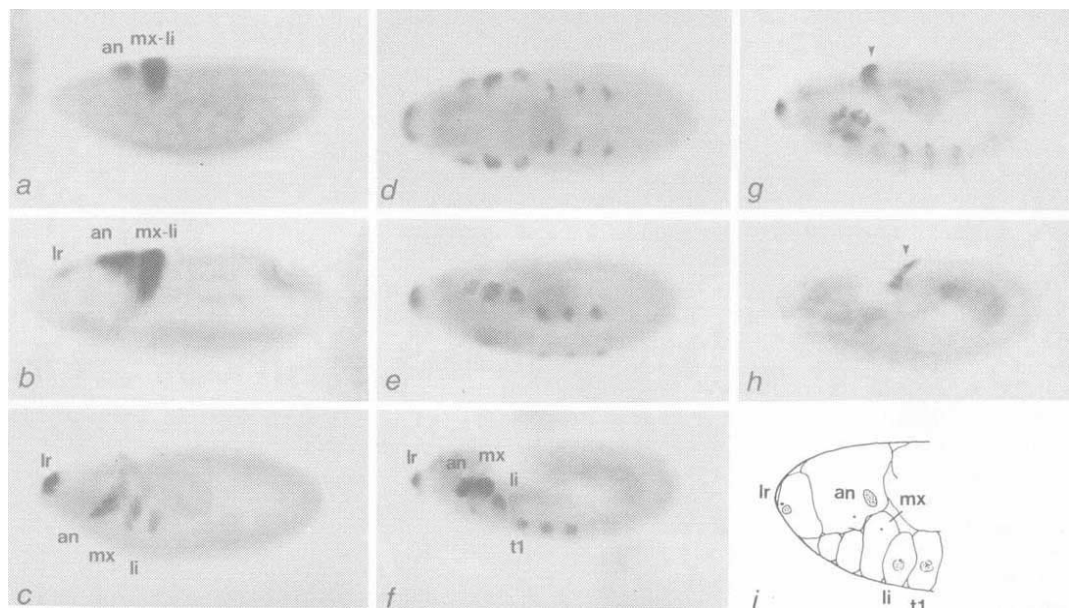
LIMB development in *Drosophila* requires the activity of a proximo-distal pattern-forming system, in addition to the antero-posterior and dorso-ventral pattern-forming systems that subdivide the embryo. Several lines of genetic evidence indicate that the *Distal-less* gene plays an important part in specifying proximo-distal positional information^{1–3}. The *Distal-less* locus encodes a homeodomain-containing protein⁴, which suggests that *Distal-less* may exert its activity through differential regulation of subordinate genes. The spatially restricted pattern of *Distal-less* expression allows direct visualization of the limb primordia during early embryogenesis. Here I report that from their inception, the leg primordia span the parasegment boundary. The segment polarity gene *wingless* seems to have a key part in defining the positions at which leg primordia will develop along the antero-posterior axis of the embryo. This analysis allows a direct molecular visualization of the compartments that subdivide the limb primordia into discrete developmental domains.

Limb primordia are believed to be established shortly after the blastoderm stage^{5–8}, by which time the embryo has been molecularly subdivided through the action of the segmentation genes^{9,10}. The antero-posterior positional information used to subdivide the embryo into parasegments also bisects the nascent primordia of the limbs into discrete developmental domains called compartments^{5–7,11}. The parasegment boundary is delineated¹² by expression of the segment polarity gene *engrailed* (*en*) in the anterior-most cells of one parasegment^{13–15} and by *wingless* (*wg*) expression in the posterior-most cells of the preceding parasegment^{16,17}. The mutually interdependent expression of *en* and *wg* is required for the stable maintenance of the parasegment through later embryonic development^{18,19}. Because the parasegment boundary seems by genetic criteria to exist in the limb primordia from their inception, and because the development of the larval legs depends critically on *wg* activity²⁰, I sought to assess the contribution of *en* and *wg* to the establishment of the limb primordia in the embryo.

The limbs of the larva include the rudimentary legs of the thoracic segments as well as structures of the gnathal and anterior head segments. Development of these structures requires *Distal-less* (*Dll*) activity^{1,2} and the pattern of *Dll* expression corresponds well to the experimentally determined locations of their primordia (ref. 21; Fig. 1f, i). *Dll* expression therefore provides an early molecular marker with which to identify the embryonic limb primordia before any overt sign of limb development. The *Dll* expression pattern is built up in a series of stages, beginning with the limb primordia of the gnathal

* Present address: Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030, USA.

FIG. 1 *Dll* expression in the limb primordia of the *Drosophila* embryo. *Dll* transcript was visualized by whole-mount *in situ* hybridization²⁹. *a*, *Dll* is first detectable during cellularization of the blastoderm in a stripe located just behind the presumptive cephalic furrow. This stripe is restricted to the dorsal third (approximately) of the embryo. Double labelling to detect *Dll* and *wg* or *en* transcripts (data not shown) indicates that the first stripe corresponds to the future maxillary and labial segments (mx-li). A second stripe appears before the completion of cellularization which corresponds to the antennal limb primordium (an). *b*, Expression in the primordium of the labrum (lr) is easily detectable shortly after the onset of germ-band extension. This expression can sometimes be faintly detected before gastrulation (data not shown). The an and mx-li signals flank the forming cephalic furrow. *c*, By the completion of germ-band elongation the maxillary (mx) and labial (li) primordia have resolved. The primordium of the labrum (lr) has migrated to the anterior tip. *d*, Ventral view showing the appearance of the primordia of the legs. Note that these primordia arise as slightly elongated clusters of cells. *e*, Vento-lateral view of a slightly older embryo showing that the clusters have rounded up. *f*, Lateral view during germ-band retraction showing the mature pattern of *Dll* expression with respect to the emerging segment borders. The leg primordium of the first thoracic segment is indicated by t1. *g*, *h*, Double labelling to detect *caudal* and *Dll* transcripts in wild-type (*g*) and a *Dll* mutant embryo (*h*) in which the *Dll* transcription unit is deleted (*Df(2R) Dll^{MP}*; ref. 4). The mutant embryo lacks the *Dll* labelling



pattern, showing that this pattern reflects *Dll* transcript accumulation. Expression of *caudal* (arrowhead) provides an independent marker for hybridization to the mutant embryo. *i*, Schematic representation of the locations of the limb primordia of an embryo at about the stage shown in *f*, determined by UV-laser-ablation fate mapping²¹. The points indicate the primordia of the larval limbs, and the stippled patches the positions of the imaginal disc primordia. The experimentally determined locations of these primordia correspond well with the domains of expression of the *Dll* transcript.

METHODS. Whole-mount *in situ* hybridization was carried out as described²⁹ with the following modification: Labelled embryos were equilibrated into 70% glycerol 30% Tris buffer (100μM), pH 7 (modified from ref. 30) for mounting, and examined using bright-field optics.

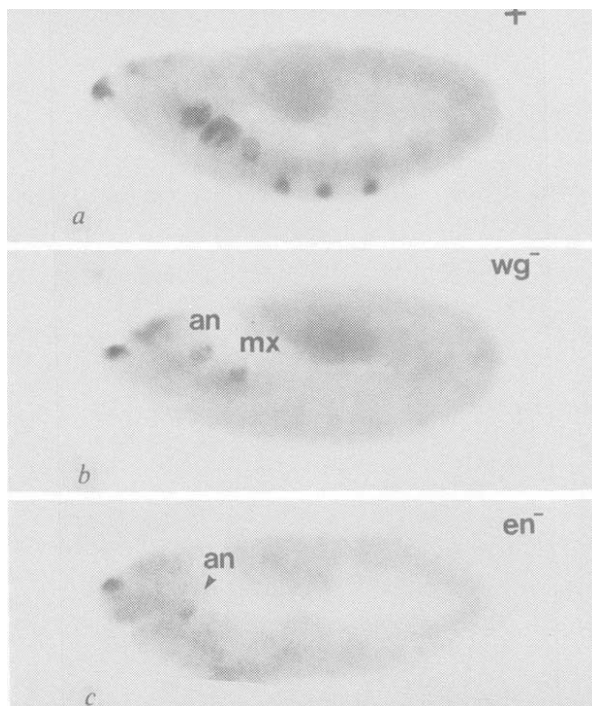


FIG. 2 Activation of *Dll* expression in the leg and labial limb primordia requires activity of both *wg* and *en*. Pattern of *Dll* expression wild-type (*a*) *wg^{c×4}* mutant (*b*) and *Df(2R) en^B* mutant (*c*) embryos. Both *wg^{c×4}* and *Df(2R) en^B* are complete lack-of-function mutations at their respective loci^{16,31}. The correspondence of the altered pattern of *Dll* expression with the genotypes of the mutant embryos was confirmed by double labelling for *Dll* and *en* transcripts (*c*) and in separate experiments (data not shown) for *wg* transcripts. In *wg* mutant embryos the primordia of the legs and labial limb do not develop. In *en* mutant embryos the primordium of the maxillary limb is also missing. Development of the labral and antennal limb primordia do not strictly depend on *wg* and *en*, indicating different requirements for activation of *Dll* in head and trunk. The genetic requirements for *en* and *wg* expression seem to differ in the gnathal segments. Both are required in the labium, and the labial limb is bisected by the parasegment boundary⁸. Expression of *en* additionally affects the maxillary and to a lesser extent the antennal limb primordia. The hypomorphic alleles *en^{IO}* and *en^{IK}* give the same result, but the morphology of the embryos is less disrupted and the strength of the signals in antenna and labrum are stronger than in the deletion mutant *en^B* (data not shown).

segments, followed by the anterior head segments and finally, shortly before germ band retraction, by the primordia of the legs (Fig. 1). *Dll* expression in the leg and labial limb primordia requires the activity of both *wg* and *en*, whereas expression in the antennal and labral limbs does not (Fig. 2). The dependence

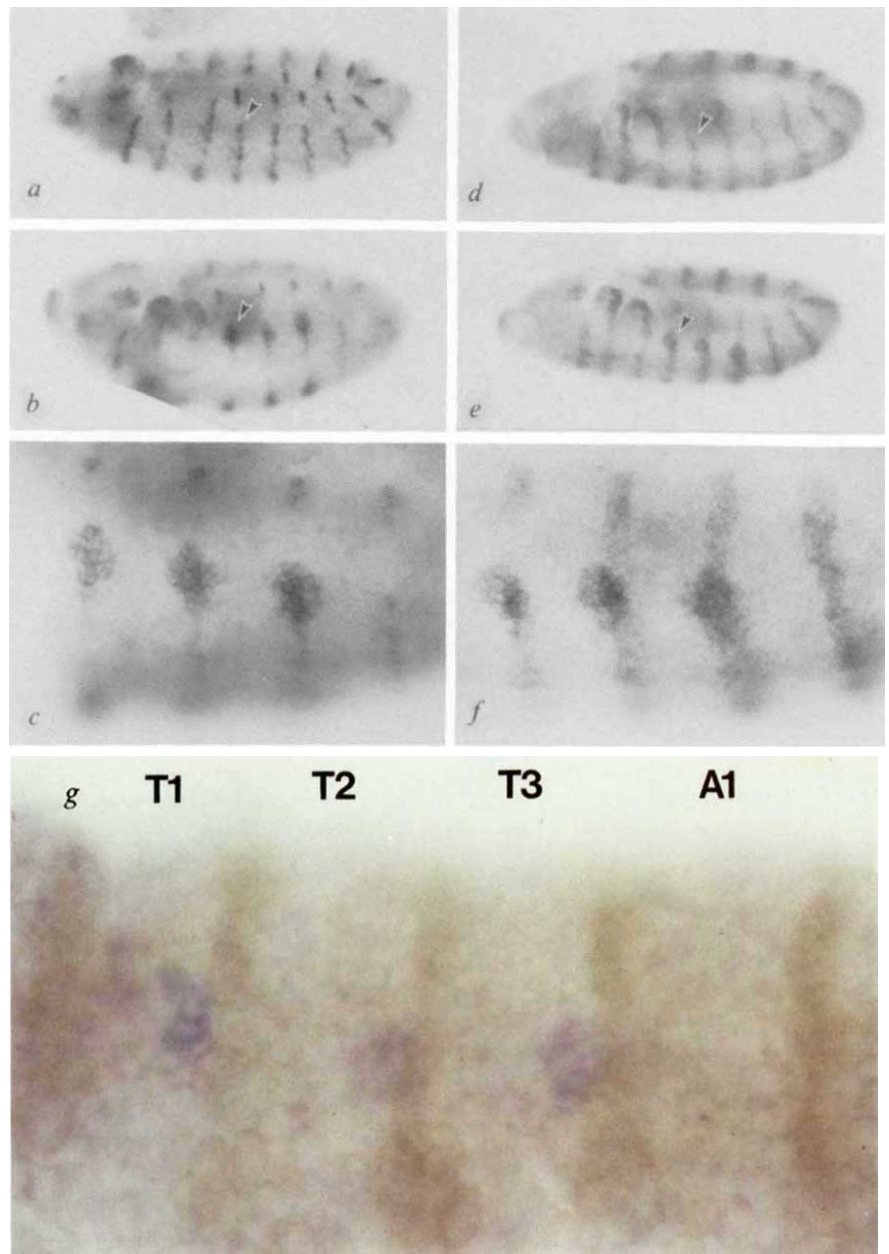
of only the labial and thoracic limbs on *wg* expression corresponds strikingly with the genetic characteristics of the adult limbs of these segments. Only the primordia of the legs and the labial limbs are bisected into compartments by the parasegment boundary during embryogenesis^{5-8,11}. No such restriction exists

FIG. 3 *Dll* expression in the leg primordia arises spanning the parasegment boundary. The spatial localization of *Dll* expression is shown with respect to *wg* and *en* expression, which delineate the parasegment boundary at the stage when *Dll* expression is first detectable. *a*, Expression of *wg*. Note that the *wg* stripes in the thoracic and abdominal segments have begun to break up laterally¹⁷. *b*, Double labelling for *wg* and *Dll* transcripts. *Dll* expression arises in a cluster of cells at the tip of the ventro-medial *wg* stripe in the thoracic segments, indicated by the arrowhead for the first thoracic segment. For comparison, the equivalent position is indicated in the *wg*-labelled embryo in *a*. *c*, Detail of the thoracic segments from another embryo labelled for *Dll* and *wg* transcripts, showing that the *Dll*-expressing cells surround the *wg*-expressing cells. Cells that express both transcripts are more strongly labelled than cells expressing either transcript alone. *Dll*-expressing cells are found close to the *wg* expressing cells, but are not necessarily in direct contact with them. This pattern is consistent with the expected functional range of the secreted *wg* gene product²⁶. *d*, Expression of *en* alone. *e*, *f*, Double labelling for *Dll* and *en* transcripts. The arrowheads indicate the location of the first thoracic leg primordium in *d* and *e*. The posterior third (approximately) of the *Dll* expressing cells also express *en*. The *Dll*-expressing and *en*-expressing cells should comprise the posterior compartment of the limb, whereas the anteriorly adjacent *Dll* expressing cells (including those that express *wg*; compare *f* with *c*) should comprise the anterior compartment of the limb. *g*, Posterior compartment of the limb is defined by cells that express both *Dll* and *en*. *Dll* transcripts are visualized as a blue reaction product in the cytoplasm of the cells of the limb primordium. The *en* gene product is visualized in the same embryo as a grey-brown stripe by antibody staining. T1, T2 and T3 indicate thoracic segments. Note that the *Dll*-expressing cells in the posterior third of the primordium also express *en*. This is best seen in T2 and T3 in this preparation. A1 indicates the first abdominal segment, which has no limb primordium. The morphology of *en*-expressing nuclei (grey-brown stripe) is poor because of the protease treatment required for *in situ* hybridization.

METHODS. Double labelling by whole-mount *in situ* hybridization and antibody staining in the same embryo was carried out with minor modification of the previously described *in situ* hybridization protocol²⁹. After standard preparation, hybridization and washing of the embryos (including protease digestion), the embryos were co-incubated for 2 h at room temperature with both monoclonal anti-*en* antibody and preabsorbed alkaline phosphatase-coupled sheep anti-digoxigenin. The embryos were washed for 4 × 20 min in PBS with 0.1% Tween-20, 2% normal goat serum and 2% normal pig serum, and then incubated for 2 h at room temperature with both biotinylated goat anti-mouse IgG (diluted 1:500) and alkaline phosphatase-coupled pig anti-sheep IgG (2 µg ml⁻¹ IgG, final concentration). Both second antibodies were pre-absorbed against fixed

at this early stage in the primordia of the adult antenna or maxilla²², consistent with the observation that these primordia develop in the absence of *wg* activity.

The leg primordia can first be visualized at the stage when the early pattern of circumferentially complete stripes of *wg* expression break up into a dorsal patch and a ventro-medial strip (ref. 17; Fig. 3*a*, *b*). Comparison of the expression patterns of *Dll* and *wg* in the same embryos shows that *Dll* expression arises in clusters of cells surrounding and including the lateral tips of the ventral *wg* stripes in the thoracic segments (Fig. 3*b*, *c*). Labelling of *Dll* and *en* transcripts shows that the *Dll* expressing cells that lie posterior to the *wg* stripe also express *en* (Fig. 3*d*, *f*). To unambiguously identify cells expressing both genes, embryos

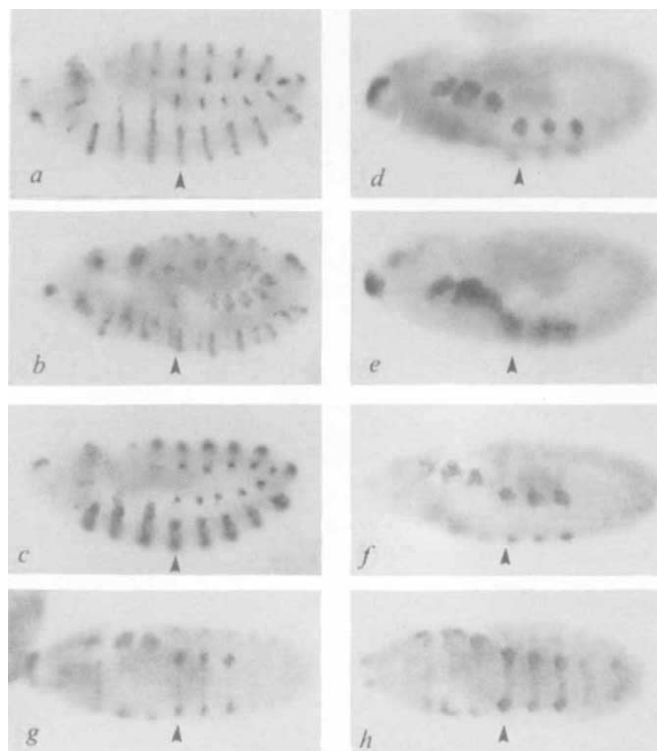


embryos at 1:10 in PBT. After washing, the embryos were treated to detect the engrailed protein under standard conditions (Vectastain elite kit, Vector labs). The embryos were then washed into the buffer of higher pH used for the alkaline phosphatase reaction and processed as previously described²⁹. Nonspecific background labelling with the antibody tends to be higher than is usual for antibody staining alone. For subsequent processing, embryos were washed into PBS (without detergent), dissected open on polylysine-coated coverslips and examined using bright-field optics.

were doubly labelled to detect *Dll* transcript and *en* protein simultaneously using different markers. Only cells in the posterior region of the cluster of *Dll*-expressing cells also express *en* protein (Fig. 3*g*). The leg primordia are therefore bisected by the parasegment boundary. Because *en* expression defines the posterior compartments of the adult limbs and of embryonic segments^{15,23} the *en*-expressing cells probably constitute the posterior compartment of the limb primordia. The *Dll*-expressing cells that do not express *en* should constitute the anterior compartment.

Comparison of the pattern of *Dll* expression with that of *wg* and *en* indicates that *wg* may provide a critical positional cue required to define the domain in which *Dll* will be activated.

FIG. 4 Ectopic expression of *Dll* follows ectopic expression of *wg* in *naked* and *patched* mutant embryos. *a-c*, Expression of *wg* in wild-type, *nkd*^{TE} and *ptc*^{IN} mutant embryos, respectively. In *nkd* mutant embryos expression of *en* expands posteriorly and an ectopic stripe of *wg* expression occurs on the opposite side of the expanded *en* stripe¹⁹, producing the doubled *wg* stripes shown in *b*. By contrast, in *ptc* mutant embryos, the *wg* stripe expands anteriorly and an ectopic stripe of *en*-expressing cells occurs¹⁸ anterior to the broadened *wg* stripe shown in *c*. *d-f*, *Dll* expression in wild-type, *nkd* and *ptc* mutant embryos, respectively. The altered pattern of *Dll* expression in the mutants follows the alteration in the *wg*-expression pattern. Embryos mutant for *nkd* can be easily identified by their abnormal *Dll* expression. Embryos mutant for *ptc* can be identified by their deep ectopic parasegmental furrows¹⁹. *g, h*, Double labelling for *Dll* and *wg* transcripts in wild-type and *ptc* mutant embryos. Note the concomitant expansion of the *wg* and *Dll* expression domains in the mutant.

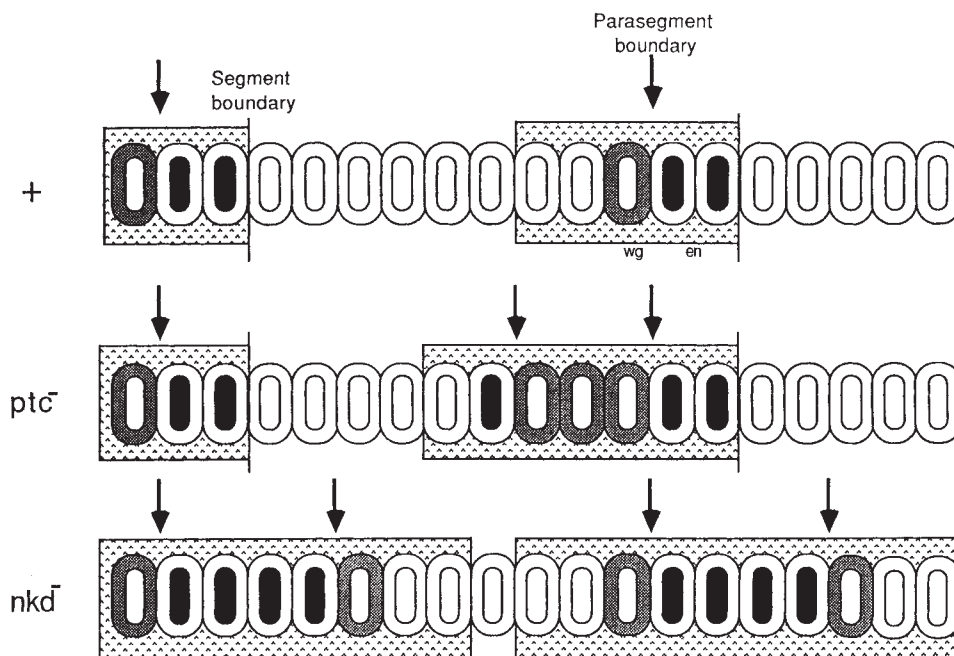


The cells that express *Dll* in the leg primordia surround and include the *wg*-expressing cells at the tip of the stripe (Fig. 3c). The *wg* gene functions nonautonomously in the embryo²⁴ and encodes a secreted product^{25,26}. Therefore, *wg* could cause localized activation of *Dll* in nearby cells. If *wg* provides a relatively direct input required for spatially localized activation of *Dll*, then the pattern of *Dll* expression should closely follow that of *wg* under aberrant conditions in mutant embryos. The late patterns of expression of both *wg* and *en* are altered in embryos mutant for the segment polarity genes *naked* (*nkd*) and *patched* (*ptc*) (refs 18 and 19; Fig. 4a-c). *Dll* expression correlates well with the changes in *wg* expression in these mutant embryos (Fig. 4d-f, h).

In *nkd* mutant embryos, *en* expression expands posteriorly forming a broadened contiguous stripe. The *wg* gene is ectopically expressed in additional stripes behind the broadened *en* stripe (ref. 19; Fig. 4b). This results in a repeated series of twice the usual number of *wg* stripes, spaced more closely than normal. In wild-type embryos, *Dll* is expressed in clusters of cells surrounding the tips of the *wg* stripes, as many as two cells away from a *wg* expressing cell (see Fig. 3c). This relationship seems

to hold true in *nkd* mutant embryos as well. Given the number and spacing of the ectopic *wg* stripes, there are very few cells in the band along the tips of the stripes that lie outside the range of *wg* influence (schematized in Fig. 5). Consequently, *Dll* expression is expanded, and in places, almost continuous in the thoracic segments (Fig. 4e). In *nkd* mutant limb primordia, the *en* stripe should be located in the centre, flanked by *wg* stripes.

FIG. 5 Schematic representation of *wg*, *en*, and *Dll* expression patterns. Cells expressing *en* are indicated by black nuclei; *wg*-expressing cells are indicated by stippled cytoplasm. The extent of *Dll* expression is indicated by the filled box in the background. At the stage when *Dll* is expressed the segments are about 12-cells wide (Figs 3c and 4d-f). One full segmental unit and the overlaps with the flanking units are shown. In wild-type embryos, the parasegment boundary (arrow) lies between the *wg*- and *en*-expressing cells. The segment boundary (line) forms behind the *en*-expressing cells. The cluster of *Dll*-expressing cells includes about two cells anterior and posterior to the *wg*-expressing cell. At later stages the limb primordium becomes separated from the segment boundary by a couple of cells. In *ptc* mutant embryos, *wg* expression expands anteriorly to form a three-cell wide stripe (refs 18 and 19; Fig. 4c) and *en* is ectopically expressed anterior to *wg* expression. If *wg* expression entrains *Dll* expression as suggested, the *Dll*-expressing zones would be expected to expand slightly anteriorly, but to remain distinctly separated. The observed pattern (Fig. 4f, h) is consistent with this prediction. In *nkd* mutant embryos, *wg* is ectopically expressed in additional stripes flanking the expanded *en* stripe (ref. 19; Fig. 4b). Both *wg* stripes would then be expected to entrain *Dll* expression in a single large zone. Because of the spacing of the *wg* 'sources', very few cells in the segment would lie outside the range of the signal, resulting in the nearly continuous band of *Dll*-expressing cells observed in *nkd* embryos (Fig. 4e). The patterns of



ectopic expression of *Dll* in the mutant embryos correlate well with the cuticular phenotypes of the mutants. In *nkd* mutant embryos, the Keilin's organs are expanded and may be fused. In *ptc* mutant embryos the Keilin's organs are expanded and may have four or five hairs, rather than the normal three, but they remain discrete and respect the normal segment boundaries (data not shown). This correspondence indicates that *Dll* could have an instructive role in organizing limb development in the embryo.

This situation alters the polarity of the primordium so that the posterior compartment lies in the centre, flanked by anterior compartments (Fig. 5). Simcox *et al.*²⁷ have observed precisely this outcome in imaginal discs derived from *nkd* mutant embryos by embryonic culture. The *nkd* mutant imaginal discs are often fused and exhibit a double anterior symmetry with the posterior compartment (defined by *en* expression) in the middle.

Dll expression also mirrors that of *wg* in *ptc* mutant embryos, in which the *wg* stripes expand anteriorly and *en* is ectopically expressed in front of the broadened *wg* stripes (refs 19 and 20; Fig. 4c). Although broader than normal, the *wg* stripes are well separated from each other in *ptc* mutant embryos (Fig. 4c). *Dll* is expressed in discrete, but broadened, patches in the mutant embryos (Fig. 4f, h) which surround the end of the *wg* stripe. In *ptc* mutant embryos, the primordia include a single *wg* stripe flanked by *en* stripes. The anterior compartment is therefore expected to lie in the centre, flanked by posterior compartments (Fig. 5). Such double posterior imaginal discs are in fact derived from *ptc* mutant embryos by *in vivo* culture²⁷. In summary, the consequences of altering *wg* expression in mutant embryos are consistent with the proposed role for *wg* as a source of information required for the activation of *Dll* expression in the leg primordia.

Dll activity is specifically required for limb development¹⁻⁴. *Dll* expression permits visualization of the nascent limb primordia before any other overt sign of limb development. In *en* and *wg* mutant embryos, which do not develop any leg structures^{20,27}, no leg primordia can be detected. In *nkd* and *ptc* mutant embryos, which show aberrant expression of *Dll*, there is a clear correspondence between the pattern of this ectopic expression and the spatial organization of both the larval and adult limbs that develop from these primordia. These observations suggest that the spatially localized activation of *Dll* expression could define the limb primordia of the embryo. This view is supported by the observation that presumptive limb cells from which *Dll* activity has been removed can contribute to the formation of the body wall, but are unable to form limb structures¹. Because *Dll* acts as a developmental switch to define the identity of cells as limb, activation of *Dll* expression could determine the limb primordia in the embryo.

Generation of the spatially restricted pattern of *Dll* expression clearly also requires some so far unidentified genetic input from the dorso-ventral pattern-forming system of the embryo. Expression of *wg* is not sufficient to activate *Dll*, because expression of the two genes is not coincident at all positions along the dorso-ventral axis of the embryo. The positional information conferred by the *wg* signal must therefore be modulated by some other dorso-ventral-restricted gene activity. There is a spatial and temporal correspondence between the change in the *wg* pattern and the activation of *Dll* expression. Both changes presumably reflect the same functional subdivision of the dorso-ventral axis of the embryo. The genes responsible for this activity have not been identified. The requirement for genetic inputs from both the antero-posterior and dorso-ventral pattern-forming systems provides a simple mechanism, for positioning the limbs accurately with respect to the other body axes and fits well with the predictions of a theoretical model proposed by Meinhardt²⁸ to explain the localized activation of limb development in the embryo. □

13. Ingham, P., Martinez-Arias, A., Lawrence, P. A. & Howard, K. *Nature* **317**, 634-636 (1985).
14. DiNardo, S., Kuner, J. M., Theis, J. & O'Farrell, P. H. *Cell*, **43**, 59-69 (1985).
15. Kornberg, T., Siden, I., O'Farrell, P. & Simon, M. *Cell* **40**, 45-53 (1985).
16. Baker, N. E. *EMBO J.* **6**, 1765-1773 (1987).
17. Baker, N. E. *Development* **103**, 289-298 (1988).
18. DiNardo, S., Sher, E., Heemskerk-Jongens, J., Kassisi, J. A. & O'Farrell, P. H. *Nature* **332**, 604-609 (1988).
19. Martinez-Arias, A., Baker, N. E. & Ingham, P. *Development* **103**, 157-170 (1988).
20. Baker, N. E. *Dev. Biol.* **125**, 96-108 (1988).
21. Jürgens, G., Lehmann, R., Schardin, M. & Nüsslein-Volhard, C. *Wilhelm Roux Arch. dev. Biol.* **195**, 359-377 (1986).
22. Morata, G. & Lawrence, P. A. *Nature*, **274**, 473-474 (1978).
23. Lawrence, P. A. & Struhl, G. *EMBO J.* **1**, 827-833 (1982).
24. Wieschaus, E. & Riggleman, R. *Cell* **49**, 177-184 (1987).
25. Rijsewijk, F. *et al. Cell* **50**, 649-657 (1987).
26. Van den Heuvel, M., Nusse, R., Johnston, P. & Lawrence, P. A. *Cell* **59**, 739-749 (1989).
27. Simcox, A. A. *et al. Development* (in the press).
28. Meinhardt, H. *Dev. Biol.* **96**, 385-387 (1983).
29. Tautz, D. & Pfeifle, C. *Chromosoma* **98**, 81-85 (1989).
30. Glicksman, M. A. & Brower, D. L. *Dev. Biol.* **126**, 219-227 (1988).
31. Gubb, D. *Wilhelm Roux Arch. dev. Biol.* **194**, 236-246 (1985).

ACKNOWLEDGEMENTS. I am indebted to Herbert Jäckle for supporting this work in his lab through a grant from the DFG Leibniz program. I thank P. Ingham, S. DiNardo and C. Nüsslein-Volhard for fly stocks; A. Simcox and P. A. Lawrence for communicating unpublished results; P. Ingham for the *wg* cDNA; P. A. Lawrence for *en* antibody; and S. DiNardo, G. Jürgens, D. Weigel, P. A. Lawrence and H. Jäckle for helpful suggestions.

Identification of a retinoic acid responsive element in the retinoic acid receptor β gene

Hugues de Thé, Maria del Mar Vivanco-Ruiz*, Pierre Tiollais, Henk Stunnenberg* & Anne Dejean

Unité de recombinaison et d'expression génétique (INSERM U 163, CNRS UA 271), Institut Pasteur, 28 rue du Dr Roux, 75724 Paris Cedex 15, France

* European Molecular Biology Laboratory, Meyerhofstrasse 1, 6900 Heidelberg, FRG

RETINOIC acid, the first morphogen described so far in vertebrates, is a vitamin A derivative which exerts striking effects on development and differentiation¹⁻³. The identification of three retinoic acid receptors as members of the nuclear receptor superfamily provides an explanation for the molecular action of morphogens on gene expression⁴⁻⁸. Functional analysis of the receptors requires the identification of target genes and of their *cis*-acting retinoic acid-responsive elements. We have previously shown that the retinoic acid receptor β gene is transcriptionally up-regulated by retinoic acid⁹ and now report the characterization of a functional retinoic acid responsive element in the β gene that mediates *trans*-activation by retinoic acid. Using deletion mapping, we have identified a 27-base pair fragment, located 59 base pairs upstream of the transcriptional start, which confers retinoic acid responsiveness on the herpes virus thymidine kinase promoter. This sequence contains a perfect direct repeat of the motif GTTCAC, which is reminiscent of the 5' half-palindrome of the thyroid and oestrogen hormone responsive elements. Specific binding of the β protein to the retinoic acid responsive element is demonstrated and is independent of the presence of retinoic acid. Both α and β receptors enhance retinoic acid response in CV1 cells, indicating that they can both act through the same DNA sequence.

To identify possible retinoic acid responsive elements (RARE) within the promoter region of the retinoic acid receptor (RAR) β gene, we isolated several human genomic clones using the 5'-most region of the RAR β complementary DNA as probe (Fig. 1a). Primer extension analysis and S1 nuclease mapping (Fig. 1c) revealed that the G indicated by +1 in Fig. 1b corresponds to the main CAP site. This conclusion is supported by the location of TATA and CAAT boxes at positions -30 and -82 respectively (Fig. 1b).

To assess whether various 5'-deleted upstream regions of the RAR β gene could direct messenger RNA transcription in a

Received 29 September; accepted 14 November 1989.

1. Cohen, S. M. & Jürgens, G. *EMBO J.* **8**, 2045-2055 (1989).
2. Sunkel, C. E. & Whittle, J. R. S. *Wilhelm Roux Arch. dev. Biol.* **196**, 124-132 (1987).
3. Cohen, S. M. & Jürgens, G. *Wilhelm Roux Arch. dev. Biol.* **198**, 157-169 (1989).
4. Cohen, S. M., Brönner, G., Küttner, F., Jürgens, G. & Jäckle, H. *Nature* **338**, 432-434 (1989).
5. Wieschaus, E. & Gehring, W. J. *Dev. Biol.* **50**, 249-263 (1976).
6. Steiner, E. *Wilhelm Roux Arch.* **180**, 9-30 (1976).
7. Lawrence, P. A. & Morata, G. *Dev. Biol.* **56**, 40-51 (1977).
8. Struhl, G. *Nature* **270**, 723-725 (1977).
9. Akam, M. *Development* **101**, 1-22 (1987).
10. Ingham, P. W. *Nature* **335**, 25-34 (1988).
11. Struhl, G. *Nature* **308**, 454-457 (1984).
12. Lawrence, P. A., Johnston, P., MacDonald, P. & Struhl, G. *Nature* **328**, 440-442 (1987).

retinoic-acid-dependent fashion, we constructed reporter plasmids by fusion of the RAR β gene promoter to the luciferase gene for transient expression in HepG2 cells (Fig. 2a). Very weak promoter activity could be detected in the absence of retinoic acid in all constructs. Addition of retinoic acid resulted in a 40-fold increase of luciferase activity for the -5 kb to +155 construct (D1 in Fig. 2a), in agreement with the magnitude of RAR β mRNA induction in these cells⁹ and in a 22-fold response for the smallest (-169 to +155) fragment, D7. Because further deletions might alter basal promoter activity, several subregions of this fragment were then fused to a thymidine kinase (TK) promoter combined to a luciferase reporter gene (TK-Luc plasmid) and transfected as before (Fig. 2b). These experiments indicate that a functional retinoic acid responsive element was present in the -62 to +1 fragment (Fig. 2b).

To analyse potential *in vitro* binding of the RAR β protein to its own promoter sequences, RAR β was overexpressed in a vaccinia expression system that yields functional proteins^{10,11}. Western blotting and immunoprecipitation of infected HeLa cell extracts revealed a protein of relative molecular mass 51,000 (51K) which was undetectable in cells infected with the wild-type virus (data not shown). Specific binding between the extracts enriched in RAR β and the -62 to +1 retinoic acid-responsive fragment was demonstrated in gel retardation experiments. A major retarded protein-DNA complex was found in extracts of vaccinia-infected cells expressing RAR β (Fig. 3a, lanes 4), but not in whole-cell extracts of wild-type virus-infected cells (lane 1). An additional band, present in most cytoplasmic or whole-cell extracts, represents a non-specific shift (Fig. 3a, b). The protein-bound probe was much more abundant in the nuclear than in the cytoplasmic fraction (lanes 4). Note that in this respect, transiently expressed RAR α and β proteins have a nuclear localization in immunocytological studies (ref. 12, and our unpublished results).

To characterize more precisely the target DNA sequence, two oligonucleotides flanking the TATA box were derived from the -62 to +1 retinoic acid-responsive fragment: O1 (-59 to -33)

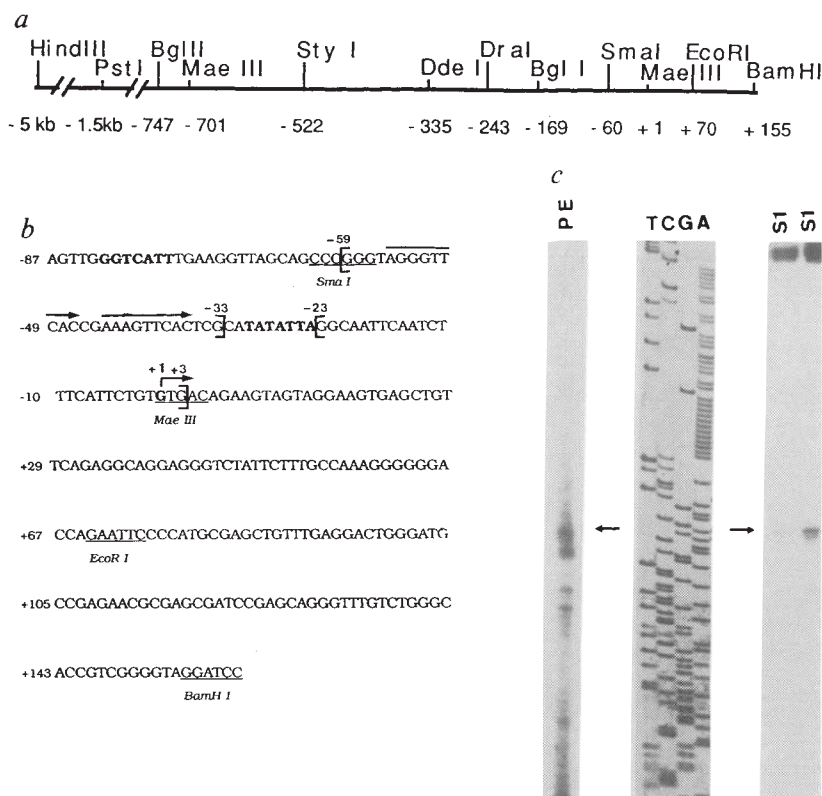
and O2 (-23 to +3) (bracketed in Fig. 1b). A retarded complex was observed with oligonucleotide O1 (Fig. 3a, lanes 5) but not with O2 (lanes 6). This reaction was specific because (1) formation of the retarded complex was not found with whole-cell extracts of cells infected with wild-type virus (lane 2), and (2) competition with a 15-fold excess of cold O1 probe caused a drastic reduction of the amount of labelled retarded complex (lanes 7), which was not observed with a 300-fold molar excess of the unrelated oligonucleotide O2 (lanes 8). This conclusion is supported by the observation that incubation with anti-RAR β anti-serum led to an upward shift of the RAR β probe complex (Fig. 3b, compare lanes 1 and 2). The antibody-induced shift could be abolished by pre-incubation of the anti-serum with the peptide used for immunization (Fig. 3b, lane 3). Taken together, these data establish that the RAR β protein can bind specifically to its own promoter region *in vitro*. The target DNA sequence, which exhibits a direct repeat of the motif A^{AA}GTTCAC, is located immediately before the TATA box (Fig. 1b). Because O1 also has a palindromic motif GGT-X₇-ACC (position -58 to -46), which is distantly related to oestrogen responsive elements¹³, we assayed a modified O1 oligonucleotide (O'1), in which the first three G residues were replaced by T, for complex retardation on gels. The retarded complexes with either O1 or O'1 were indistinguishable (data not shown), confirming that this palindromic structure is not responsible for the binding of RAR β to O1.

We found that retinoic acid administration has no effect on the formation of the retarded complexes (Fig. 3b, compare lanes 4 and 5). Similar results have been obtained using the triiodothyronine (T₃) receptor (ref. 14; and J. Sap and H.S., unpublished data). To establish whether this is relevant to the *in vivo* situation requires further investigation.

To demonstrate that O1 constitutes a functional retinoic acid responsive element, three copies of the 27-mer were inserted into TK-Luc to generate RARE₃-TK-Luc. Retinoic acid administration elicited a 40-fold induction of luciferase activity in HepG2 cells transfected with this plasmid (Fig. 2c), showing

FIG. 1 a, Restriction map of the RAR β 5'-flanking region. Only restriction sites used in reporter gene constructs have been indicated upstream of the Bgl/II site. Nucleotide positions are numbered relative to the CAP site. b, Nucleotide sequence of the RAR β promoter. The TATA and CAAT boxes are shown in bold characters. The main CAP site is indicated by an arrow. The two oligonucleotides used in gel-retardation experiments are bracketed and the direct repeat in the retinoic acid-responsive element is marked by two arrows overhead. c, Mapping of the RAR β mRNA CAP site: primer extension (PE) and S1 nuclease analysis (S1) of RAR β mRNA; the lanes marked TCGA show dideoxy sequencing of a recombinant M13 clone which was used as a size marker. The arrows show the position of the 5' end of the RAR β mRNA.

METHODS. The genomic library was generated by inserting partial *Sau*III fragments of human DNA in the *Bam*HI site of λ EMBL3 and screened with a *Bam*HI-*Nci*I fragment (position 8 to 88 in the RAR β cDNA; ref. 6) as a probe. The region extending from the Bgl/II site to the *Bam*HI site (position 8 in ref. 6) was sequenced by the dideoxy chain-termination procedure¹⁸. Primer extension has been described¹⁹: a [γ -³²P]ATP kinased oligonucleotide (5'-ATCCTACCCGACGGTGCC3', complementary to the region +139 to +158 of the RAR β promoter) was used as a primer with 50 μ g RNA from retinoic acid-induced PLC/PRF/5 cells. The 217-base pair single-stranded probe used for S1 protection was generated from the same 5' end-labelled oligonucleotide annealed to the single-stranded *Bam*HI-*Pst*I promoter fragment, elongated with Klenow polymerase, digested with *Sma*I and isolated on a sequencing gel. Hybridization and digestion with nuclease S1 were as



described¹⁹, using 50 μ g RNA from non-induced (left lane) and retinoic acid-induced (right lane) PLC/PRF/5 cells.

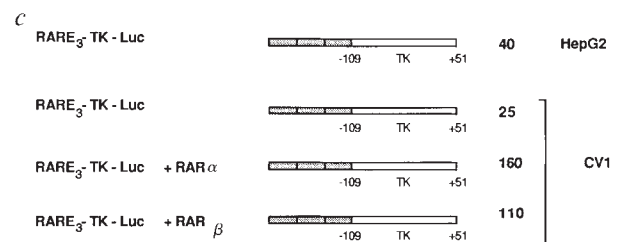
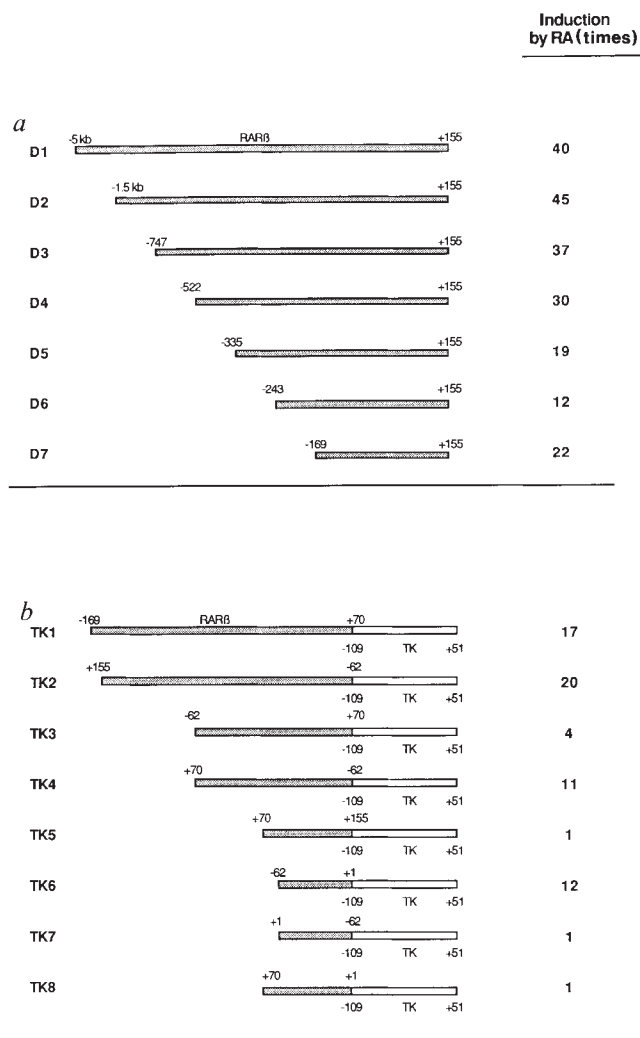


FIG. 2 *a*, Deletion mapping of the RARE. Fragments of the *RARβ* promoter were fused to the luciferase cDNA and SV40 polyadenylation signal. Restriction sites used to generate these fragments are indicated by their positions. The resulting constructs were transiently expressed in HepG2 cells and assayed for responsiveness to retinoic acid. *b*, Transfer of retinoic acid regulation to the TK promoter. *RARβ* gene fragments were inserted in a reporter TK-Luc plasmid and assayed as before. *c*, A 27-mer oligonucleotide (–59 to +33) containing the RARE was inserted in the TK-Luc construct yielding RARE₃-TK-Luc. This reporter transfected HepG2 cells or cotransfected CV1 cells with the *RARα* (pSG5RARα; ref. 8) or *RARβ* (RSV-RARβ; H.deT., unpublished) expression vectors.

METHODS. The luciferase cDNA and SV40 polyadenylation site were transferred from pSVOAL (ref. 20) to pTZ18, and *RARβ* 5' sequences were inserted upstream from the luciferase ATG. TK-Luc plasmid was obtained by replacing the chloramphenicol acetyl transferase (CAT) gene in pTK_{cat}-CAT, a CAT expression vector containing a TK promoter fragment (–109 to +51) (P. Herbolmel, unpublished) by the luciferase cDNA from pSVOALΔ5' (ref. 20). *RARβ* gene fragments were inserted at the *Sa*I site (position –109) of the TK promoter. Three copies of the O1 oligonucleotide, 5'GGGTAGGGTTCACCGAAAGTTCACCTCG3', with a 5' TCGA overhang, were inserted at the same *Sa*I site to generate RARE₃-TL-Luc. The reporter plasmids (10 μg), the β-galactosidase expression vector pCH110 (10 μg) and, when appropriate, the *RARα* or *RARβ* expression vectors or the Rous sarcoma virus-globin control plasmid (2 μg) were introduced into HepG2 or CV1 cells (1.5 × 10⁶ per 25 cm² dish) by calcium phosphate co-precipitation. The cells were maintained in 10% dextran/charcoal-stripped (for HepG2 cells) or de-lipidified (for CV1 cells) serum for 24 h, with or without 10^{–6} M retinoic acid. Luciferase expression was quantified²¹ in duplicate experiments and normalized for β-galactosidase activity. Induction was estimated as the mean of the ratio between retinoic acid-induced and non-induced specific luciferase activity from several experiments.

FIG. 3 DNA-binding of *RARβ* to its own promoter sequence *in vitro*. *a*, Identification of the target DNA sequence in the *RARβ* gene. Cytoplasmic (C) and nuclear (N) extracts of recombinant vaccinia virus-infected HeLa cells were incubated with the ³²P-labelled –59 to +1 fragment (lanes 4), the –59 to –33 oligonucleotide (O1) (lanes 5) and the –23 to +3 oligonucleotide (O2) (lanes 6). As a control, whole-cell extracts (W) from wild-type virus-infected cells were incubated with the same probes (lanes 1–3). Protein-bound and free DNA were separated by gel electrophoresis. Competition of the binding of O1 to infected cell extracts was performed using a 15-fold excess of specific unlabelled O1 DNA (lanes 7) or a 300-fold excess of non-specific cold O2 DNA (lanes 8). *b*, Effects of anti-*RARβ* anti-serum or retinoic acid on the retarded complex. *RARβ*-enriched nuclear extracts were incubated with probe O1 in the presence of pre-immune serum (lane 1), anti-*RARβ* anti-serum (lane 2) and anti-*RARβ* plus P17 peptide (lane 3). Influence of retinoic acid added *in vitro* on the formation of the DNA-protein complex: nuclear or cytoplasmic extracts of recombinant virus-infected cells grown in de-lipidified medium were incubated with the O1 probe without (lanes 4) or with (lanes 5) added retinoic acid.

METHODS. HeLa cells were infected with a recombinant vaccinia virus carrying the *RARβ* cDNA under the transcriptional control of the late promoter of the 11K-protein gene. Cytoplasmic or nuclear extracts corresponding to 10⁵ cells were incubated in (final concentrations) 7.5% glycerol, 0.05 mM EDTA, 10 HEPES, pH 7.9, 5.5 mM MgCl₂, 25 mM KCl, 3 mM spermidine and

0.1 mg ml^{–1} poly d(I-C), with 40,000 c.p.m. (90 fmol) of a ³²P-labelled probe; the reaction mixture was kept for 1 h at 0°C and loaded onto a 5% polyacrylamide gel. The anti-*RARβ* anti-serum was raised in rabbits against the synthetic peptide P17 (ValArgAsnAspArgAsnLysLysLysLysGluThrSerLys-GlnGluCys, corresponding to amino acids 151–167; ref. 6) (Appligene, Strasbourg) coupled to keyhole limpet haemocyanin (KLH) by monthly immunizations of 300 μg KLH-coupled peptide emulsified in complete (and subsequently incomplete) Freund's adjuvant. For the experiments involving anti-serum, 1 μl pre-immune or immune serum and, when appropriate, 1 μg P17 peptide were added to the 20 μl binding reaction.

that this sequence is not only sufficient to confer retinoic acid responsiveness, but is also highly efficient. Luciferase mRNAs were analysed by S1-nuclease mapping to ensure that retinoic acid treatment resulted in an increase of correctly initiated transcripts (data not shown). In this experiment, endogenous HepG2 RARs were responsible for the retinoic acid-dependent *trans*-activation. To test whether RAR α and/or β could both activate transcription of this reporter, CV1 cells were chosen for cotransfection with RAR expression vectors because of their low RAR content¹⁵. A 25-fold increase in luciferase activity was observed in the absence of RAR α or β co-expression, indicating the presence of functional endogenous RARs. But when RAR α and β expression vectors were used for cotransfection, retinoic acid elicited a significantly greater response (160- and 110-fold, respectively; Fig. 2c), indicating that both RAR α and β can recognize the same RARE and mediate *trans*-activation of the RAR β gene.

We have characterized a natural RARE which is present in the RAR β promoter region and which confers retinoic acid inducibility on the RAR β gene. The responsive element, located within the 27-base pair fragment immediately adjacent to the TATA box, binds the RAR β protein *in vitro* and can confer retinoic acid regulation on a heterologous promoter. This RARE contains a perfect direct repeat of the motif GTTCAC, which is related to the GGTCA half-palindrome of thyroid and oestrogen hormone responsive elements¹³. The presence of a directly repeated sequence is puzzling, in view of the fact that other hormone responsive elements exhibit dyad symmetry^{13,16}. It will be interesting to define the nature of the protein complex that interacts with this retinoic acid-responsive element.

Both RAR α and β mediate retinoic acid inducibility through the natural RARE in transient expression experiments, although enhancement of the response is weak. Induction of the endogenous RAR β in retinoic acid-treated cells, masking the effect of transfected RAR expression vectors, could partly account for this observation. We have previously shown that the RAR β gene transcript is up-regulated by retinoic acid *in vivo*⁹ and now report that the RAR β protein can bind *in vitro* to its own promoter sequence. It is not yet clear whether or not these results reflect the existence of a positive autoregulatory loop.

A thyroid responsive element is known to confer retinoic acid inducibility on the TK promoter in gene transfer experiments¹⁵, and thyroid hormone receptor expression interferes with retinoic acid response through this element¹⁷. The availability of a RARE which is poorly related to canonical thyroid responsive elements could be useful in determining the extent to which the retinoic acid and thyroid hormone networks of target genes overlap.

Very mild muscular dystrophy associated with the deletion of 46% of dystrophin

S. B. England*, L. V. B. Nicholson†, M. A. Johnson†, S. M. Forrest*, D. R. Love*, E. E. Zubrzycka-Gaarn‡, D. E. Bulman‡, J. B. Harris§ & K. E. Davies*||

* Molecular Genetics Group, Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford, OX3 9DU, UK

† Muscular Dystrophy Group Research Laboratories, Regional Neurological Centre, Newcastle General Hospital, Westgate Road, Newcastle upon Tyne, NE4 6BE, UK

‡ Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, MSG 1X8, Canada

DUCHENNE muscular dystrophy (DMD)¹ and Becker muscular dystrophy (BMD), a much milder form of the disease where the age of onset can sometimes be as late as the third or fourth decade of life, are caused by mutations in the same X-linked gene²⁻⁷, a 14 kilobase (kb) transcript^{6,7} which is spread over more than 2 megabases of the human X chromosome⁸⁻¹⁰. The corresponding protein, dystrophin, has a relative molecular mass of 400,000 (ref. 11). Most mutations causing DMD and BMD are deletions^{7,12,13} (reviewed in ref. 14) and deletions associated with both phenotypes are observed throughout the gene sequence. This observation led to the suggestion¹⁵ that DMD patients possess deletions that disrupt the reading frame of the protein, whereas BMD patients have deletions that retain the translational reading frame and enable the muscle cells to produce altered dystrophin products. This theory is supported by immunoblotting studies, which show that DMD patients lack dystrophin in their muscle cells or that dystrophin is present at very low levels, whereas BMD patients produce a protein with reduced abundance or abnormal size¹⁶. Here we describe a deletion of the dystrophin gene in a family segregating for very mild BMD, one member of which was still ambulant at age 61 years, which removes a central part of the dystrophin gene encompassing 5,106 base pairs of coding sequence, almost half the coding information. Immunological analysis of muscle from one of the patients demonstrates that this mutation results in the production of a truncated polypeptide localized correctly in the muscle cell. These results are particularly significant in the context of gene therapy which, if it is ever envisaged, would be facilitated by the replacement of the very large dystrophin gene with a more manipulatable mini-gene construct.

The proband (IV-5; patient 1353) in the BMD family shown in Fig. 1b presented to the neurologist because of a family history of muscle disease. A male relative (III-1; patient 324) showed symptoms of muscle weakness in his thirties, but the progression of his disease was slow enough that he could still walk with the aid of a stick at age 61 years. A second male relative (II-2) died of a cerebral haemorrhage at the age of 52 years but was walking with the aid of a stick prior to his death (Fig. 1b; see ref. 17 for a full clinical description). We analysed the mutation in this family using complementary DNA probes from the DMD gene^{12,18}; the results are given in Figs 1 and 2. We performed *Pst* I and *Hind*III digests so that our results could be correlated with the exon order^{7,13,19}. The 5' end of the deletion present in these patients has been previously shown to be in the region of pERT87-15 (ref. 17) and hence in the distal 3' portion of cDNA Cf27 (ref. 18). Hybridization of genomic DNA with the 3' end of Cf27 led to the identification of two *Hind*III fragments of 6 kb and 1.7 kb, which correspond to exons 16 and 17, respectively. In the case of affected patients

Received 9 October; accepted 20 November 1989.

- Strickland, S. & Mahdavi, M. *Cell* **15**, 393-403 (1978).
- Thaller, C. & Eichele, G. *Nature* **327**, 625-628 (1987).
- Roberts, A. & Sporn, M. in *The Retinoids* (eds Sporn, M., Roberts, A. & Goodman, D.) 209-286 (Academic, Orlando, 1984).
- Giguere, V., Ong, E. S., Segui, P. & Evans, R. M. *Nature* **330**, 624-629 (1987).
- Petkovich, M., Brand, N. J., Krust, A. & Chambon, P. *Nature* **330**, 444-450 (1987).
- de Thé, H., Marchio, A., Tiollais, P. & Dejean, A. *Nature* **330**, 667-670 (1987).
- Brand, N. *et al.* *Nature* **332**, 850-853 (1988).
- Zelent, A., Krust, A., Petkovich, M., Kastner, P. & Chambon, P. *Nature* **339**, 714-717 (1989).
- de Thé, H., Marchio, A., Tiollais, P. & Dejean, A. *EMBO J.* **8**, 429-433 (1989).
- Sap, J., Munoz, A., Schmitt, J., Stunnenberg, H. & Vennström, B. *Nature* **340**, 242-244 (1989).
- Schmid, W., Strähle, U., Schütz, G., Schmitt, J. & Stunnenberg, H. *EMBO J.* **8**, 2257-2263 (1989).
- Gaub, M. P. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **86**, 3089-3093 (1989).
- Beato, M. *Cell* **56**, 335-344 (1989).
- Damm, K., Thompson, C. & Evans, R. *Nature* **339**, 593-597 (1989).
- Umesono, K., Giguere, V., Glass, C., Rosenfeld, M. & Evans, R. *Nature* **336**, 262-265 (1988).
- Evans, R. *Science* **240**, 889-895 (1988).
- Graupner, G., Willis, K., Zukerman, M., Zhang, X. & Pfahl, M. *Nature* **340**, 653-656 (1989).
- Sanger, F., Nicklen, S. & Coulson, A. K. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
- Greene, J. & Struhl, K. in *Current Protocols in Molecular Biology* Vol. 1 (eds Ausubel, F. M. *et al.*) 4-6 (Wiley, New York, 1987).
- de Wet, J., Wood, K., de Luca, M., Helinsky, D. & Subramani, S. *Molec. cell. Biol.* **7**, 725-737 (1987).
- Nguyen, V., Morange, M. & Bensaude, O. *Analyt. Biochem.* **171**, 404-408 (1988).

ACKNOWLEDGEMENTS. We thank M. Yaniv and J. Weissbach for advice, discussion and comments on the manuscript. S. Subramani for pSV0AL and pSV0AL Δ 5', P. Chambon for pSG5RAR α , P. Herbolme for pRSVgI and pTK α -CAT, O. Bensaude for advice on the luciferase assay, A. Marchio for assistance and S. Cole for revising the manuscript. This work was supported by grants from the Association pour la Recherche contre le Cancer, la Ligue Nationale Française contre le Cancer, and the National Institutes of Health.

§ Present address: The Murdoch Institute, Flemington Road, Parkville, Victoria 3052, Australia.

|| To whom correspondence should be addressed.

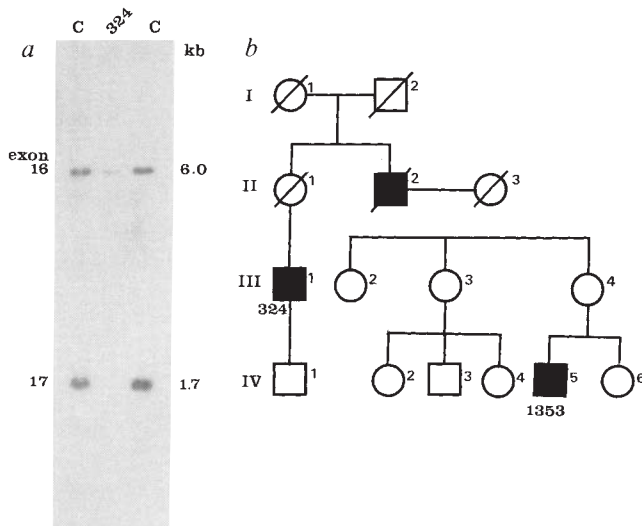


FIG. 1 *Hind*III digest of patient 324 and controls (c) hybridized to a Cf27 derivative detecting exons 16 and 17 (ref.18). The Southern blot was performed as described previously¹² and washed in 0.1×SSC at 65°C. Autoradiography was at -70°C for 3 days.

(1353 and 324), however, exon 17 is deleted. The proximal breakpoint is shown in Fig. 1a.

The analysis of the distal breakpoint of the deletion in this family is shown in Fig. 2. *Hind*III fragments of 3.9 kb and 10 kb which correspond to exons 47 and 48, respectively, are missing in patient 324, although a clear band is seen at 1.6 kb corresponding to exon 49 (Fig. 2a). Because of the similarity of the lengths of the *Hind*III genomic fragments containing exons 48 and 50, we confirmed the position of the distal breakpoint in a *Pst*I digest. As seen in Fig. 2b, the deletion encompasses *Pst*I bands at 5.4 kb and 5.1 kb, corresponding to exons 47 and 48. The band at 6.8 kb corresponding to exon 49 is clearly visible. Thus the deletion in this family includes exons 17-48 and removes 5,106 bp of coding sequence. Although we analysed the patient's DNA with 12 different restriction enzymes using cDNA probes from both sides of the deletion, no changed band was detected. It is highly unlikely, therefore, that the breaks are in or very near the exons. According to the analysis of Koenig *et al.*²⁰, this deletion should not disrupt the translation reading frame of dystrophin. From the physical map of the gene constructed by pulsed field gel electrophoresis and analysis of patient 324, we estimate that the deletion encompasses approximately 700 kb of genomic DNA (ref. 21; S.B.E., unpublished observations).

Our analysis indicates that the affected patients should produce a truncated dystrophin molecule of relative molecular mass about 200,000 ($M_r \sim 200K$). An open muscle biopsy (m. quadriceps femoris) of patient 1353 was therefore taken to determine

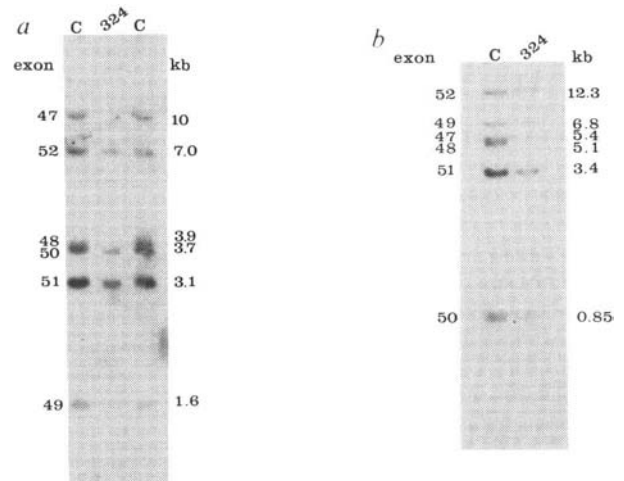
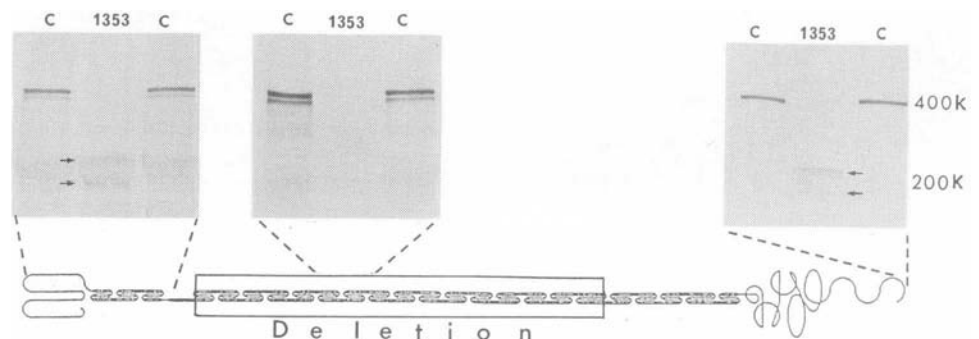


FIG. 2 A *Hind*III digest of patient 324 DNA and controls (c) hybridized to Cf56a (the second band detected by exon 48 at 1.2 kb is not shown in this gel). The Southern blot was performed as described previously¹² and washed at 0.1×SSC at 65°C. Autoradiography was at -70°C for 2 days. b, *Pst*I digest of patient 324 DNA and controls (c) hybridized to Cf56a.

the size and location of his dystrophin protein. In contrast to his mild clinical phenotype, a histopathological examination of the patient showed severe muscle fibre atrophy and extensive replacement by fat and fibrous connective tissue. Surviving fibres of normal diameter showing compensatory hypertrophy were confined to only a few areas of the large tissue sample. Thus the histopathological features were consistent with diagnosis of BMD but seemed much more severe than might be expected from the clinical findings. Patient 1353 is now 25 years of age, with the upper body of a weight trainer (which he is) and it is possible therefore, that the biopsy obtained from m. quadriceps might be unrepresentative of the patient's musculature in general.

Protein was extracted from a portion of the biopsy sample and analysed by western blotting. The blot was probed with antibodies derived from the 5' end, central part and the 3' end of dystrophin. Figure 3 shows that the muscle from the patient gives a positive signal with antibodies raised against the two ends of the protein but a negative signal with the antibody raised against the central domain. The positively labelled polypeptide has an $M_r \sim 200K$. The band appears as a doublet, which is probably the result of an artefactual separation due to the high level of myosin in this position of the gel. The myosin band was identified in parallel experiments with an anti-myosin monoclonal antibody (data not shown). Although the apparent relative intensity of the two bands appears to vary between the N- and C-terminal antibodies, this is an artefact created by the

FIG. 3 Western blot analysis of dystrophin in muscle taken from patient 1353. SDS-polyacrylamide gel electrophoresis and blotting were performed as described previously²². Normal control muscle (c) and muscle from patient 1353 were homogenized in 20 volumes of a sample treatment buffer containing 15% SDS. The control tissue extract was then diluted fivefold so that approximately the same amount of 'muscle' protein was applied to the gel from each sample (40 µg in 30 µl). Duplicate blots were then incubated with different primary antibodies and horseradish peroxidase conjugated second antibodies. The 5' antibody (9219) is a sheep polyclonal antibody



(Ab), the central antibody (Dy4/6D3) is a mouse monoclonal Ab and the 3' antibody (1460) is an affinity purified rabbit polyclonal Ab.

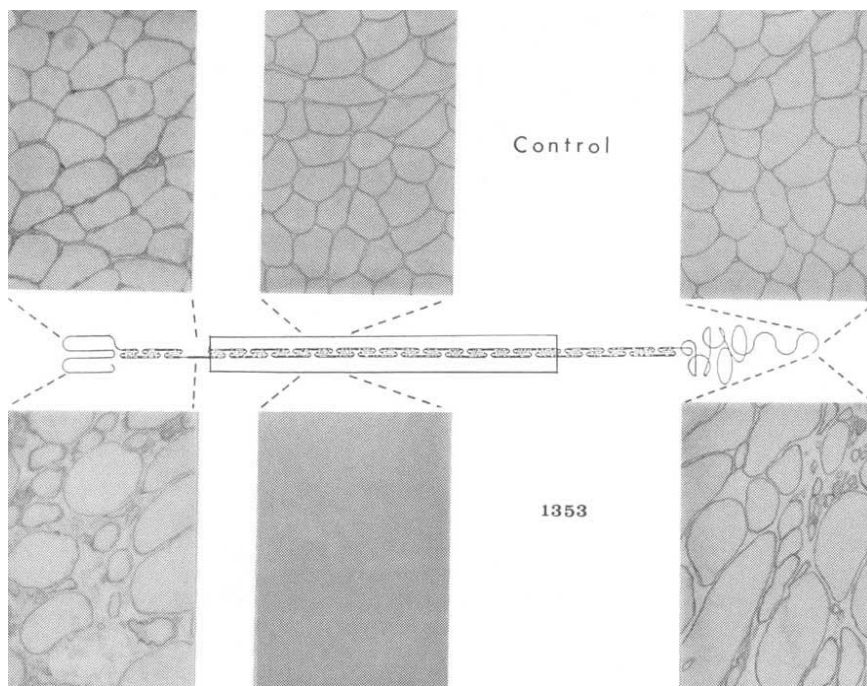


FIG. 4 Immunocytochemical analysis of dystrophin in unfixed frozen tissue sections from the biopsy of patient 1353 and from a normal control subject. The primary and secondary antibodies are as described in the legend for Fig. 3.

separate photography of the blots. These data are consistent with the DNA deletion described above and indicate that these patients do indeed produce a truncated dystrophin molecule.

Immunocytochemistry was conducted as described previously²³ and the results are shown in Fig. 4. Positively staining muscle fibres can be seen with antibodies raised against the 5' and 3' ends of the dystrophin molecule, but no immunolabelling is seen with the antibody raised against the central domain. This analysis confirms that the affected individuals in this family produce a truncated dystrophin and that it is localized correctly in the muscle cell^{24,25}.

The data presented here show that even very large deletions of the dystrophin gene can give rise to a mild phenotype. The patients described produce a protein which is approximately half of the normal size. Generally, a mild phenotype could be explained by compensation of the defect by other dystrophin-related polypeptides²⁶ or by the partial function of the truncated molecule. The mild phenotype breeds true in this family and therefore must be related to the locus at Xp21. The conformation

of the smaller protein must permit almost normal interactions with the other components of the muscle cytoskeleton.

The deletion in the patients presented here occurs in the spectrin-like repeat structure of the molecule²⁷. The domains thought to be involved in interactions with other muscle proteins, such as actin, remain intact. The 5' end of the deletion at the exon 16/17 boundary corresponds precisely to the end of a repeat. At the 3' end of the deletion the repeat finishes 5 residues after the exon boundary, but this extra sequence is unlikely to affect the folding of the molecule. Thus this deletion may simply shorten the dystrophin molecule. More detailed studies of this truncated molecule will be required to elucidate why it functions so well.

One of the difficulties in replacing dystrophin, either for functional studies or in gene therapy experiments, is that the full 14-kb sequence is too large for the common vectors. The 'minigene' which functions so effectively in these patients should allow us to develop a good understanding of the 3' and 5' domains of dystrophin and assist in the development of ways to treat the disease in the future. □

- Emery, A. E. H. *Duchenne Muscular Dystrophy* (eds. Harper, P. & Bobrow, M) (Oxford University Press, 1987).
- Davies, K. E. *et al. Nucleic Acids Res.* **11**, 2302-2312 (1983).
- Kingston, H. M. *et al. Lancet* **2**, 1200 (1983).
- Kunkel, L. M. *et al. Nature* **322**, 73-77 (1986).
- Monaco, A. P. *et al. Nature* **316**, 842-845 (1985).
- Ray, P. N. *et al. Nature* **318**, 672-675 (1985).
- Koenig, M. *et al. Cell* **50**, 509-517 (1987).
- Burmeister, M. & Lehrach, H. *Nature* **324**, 582-585 (1986).
- van Ommen, G.-J. B. *et al. Cell* **47**, 499-504 (1986).
- Kenwick, S. *et al. Cell* **48**, 351-357 (1987).
- Hoffman, E. P., Brown, R. H. & Kunkel, L. M. *Cell* **51**, 919-928 (1987).
- Forrest, S. M. *et al. Nature* **329**, 638-640 (1987).
- Darras, B. T. & Francke, U. *Am. J. hum. Genet.* **43**, 612-619 (1988).
- Love, D. R. & Davies, K. E. *Molec. biol. Med.* **6**, 7-17 (1989).
- Monaco, A. P. *et al. Genomics* **2**, 90-95 (1988).
- Hoffman, E. P. *et al. N. Engl. J. Med.* **318**, 1363-1368 (1988).
- Davies, K. E. *et al. J. Med. Genet.* **99**, 99 (1987).

- Smith, T. J., Forrest, S. M., Cross, G. S. & Davies, K. E. *Nucleic Acids Res.* **15**, 9761-9769 (1987).
- Forrest, S. M. *et al. Genomics* **2**, 109-114 (1988).
- Koenig, M. *et al. Am. J. hum. Genet.* **45**, 498-506 (1989).
- Den Dunnen, J. T. *Am. J. hum. Genet.* in the press.
- Nicholson, L. V. B. *et al. J. neurol. Sci.* **94**, 125-136 (1989).
- Nicholson, L. V. B. *et al. J. neurol. Sci.* **94**, 137-146 (1989).
- Zubrzycka-Gaarn, E. E. *et al. Nature* **333**, 466-469 (1988).
- Watkins, S. C., Hoffman, E. P., Slayter, H. S. & Kunkel, L. M. *Nature* **333**, 863-866 (1988).
- Love, D. R. *et al. Nature* **339**, 55-58 (1989).
- Davison, M. D. & Critchley, D. R. *Cell* **52**, 159-160 (1988).

ACKNOWLEDGEMENTS. We thank Professor Adrian Williams and Dr Sarah Bunday for access to the family. We thank Tracey Flint, Rosalind Marsden, Keith Davison and Martin Barron for technical assistance and Helen Blaber for typing the manuscript. We are grateful to Terry Smith for helpful discussions. We are indebted to the Muscular Dystrophy Group of Great Britain and Northern Ireland, the Muscular Dystrophy Association, U.S.A. and the Medical Research Council of Great Britain for their support. S.B.E. is a recipient of an ORS award.

A transgenic mouse model of sickle cell disorder

David R. Greaves, Peter Fraser, Miguel A. Vidal, Michael J. Hedges*, David Ropers*, Lucio Luzzatto* & Frank Grosfeld

Laboratory of Gene Structure and Expression, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK
* Haematology Department, Royal Postgraduate Medical School, Hammersmith Hospital, Du Cane Road, London W12 0HS, UK

A SINGLE base-pair mutation (β^s) in codon 6 of the human β -globin gene, causing a single amino-acid substitution, is the cause of sickle cell anaemia¹. The mutant haemoglobin molecule, HbS, polymerizes when deoxygenated and causes deformation of the erythrocytes to a characteristic 'sickled' shape. Sickling of cells in small vessels causes painful crises and other life-threatening complications^{2,3}. Although the molecular basis for sickle cell anaemia has been known for 30 years, no definitive treatment is available⁴. An animal model of sickle cell anaemia would not only allow a detailed analysis of the factors that initiate erythrocyte sickling *in vivo* and of the pathophysiology of the disease, but would also permit the development of novel approaches to the treatment of the disease. By using the dominant control region sequences from the human β -globin locus, together with human α - and β^s -globin genes, we have obtained three transgenic mice with HbS levels ranging from 10 to 80% of total haemoglobin in their red cells. As observed in homozygous and heterozygous Hb^s patients, the erythrocytes of this mouse sickle readily on deoxygenation. Irreversibly sickled cells^{2,3}, which are characteristic of sickle-cell patients homozygous for β^s , are also observed in the peripheral blood of the mouse with high levels of HbS.

The dominant control region (DCR) sequences, which flank the human β -globin locus, direct high-level copy-number-dependent expression of the human β -globin gene in erythroid cells and in transgenic mice^{5,9}. When these β -globin DCR sequences are linked to the human α -globin gene in transgenic mice, high-level expression is obtained⁷⁻¹⁰. It was therefore possible to obtain transgenic mouse lines expressing more human haemoglobin (HbA) than endogenous mouse haemoglobin by using a construct containing the β -globin DCR, the human β - and the α_1 -globin gene⁸. These mice had a normal mean corpuscular volume (MCV) and mean corpuscular haemoglobin concentration (MCHC)⁸, but showed an imbalance in the synthesis of α - and β -globin, as only one α -gene, rather than the normal two α -genes, was used⁸. Our initial constructs for expressing HbS in transgenic mice included two human α -globin genes and the human β^s -globin gene (Fig. 1).

We obtained five transgenic animals and determined the transgene copy number by Southern blot analysis of tail-biopsy DNA (Fig. 2a, and data not shown). These mice fall into the usual two categories—mice that express at low levels because of deletions or mosaicism, and mice that are fully transgenic and

express at high levels^{5,7,8}. Mouse 41 had one copy of the transgene but had deletions in the β -globin gene (Fig. 2a) and the DCR (data not shown). The other transgenic mice did not have any detectable rearrangements and had the copy numbers shown in Table 1. We analysed RNA prepared from peripheral blood samples by S1 nuclease protection to determine the level of expression of the transgenic globin genes. Mouse 41 has a low level of correctly initiated β^s -globin messenger RNA (5% of mouse β -globin mRNA) caused by the deletion in the transgene. Two of the mice are mosaic and have low (mouse 75) or undetectable (mouse 64) levels of β^s -globin mRNA. Two mice have high levels of human α - and β^s -globin mRNA in peripheral blood and show, after correction for specific activities of the probes, a near-balanced synthesis of human α - and β^s -globin mRNA (Fig. 2c and Table 1). We analysed peripheral blood haemolysates by isoelectric focusing under conditions that resolve human haemoglobin (HbA), sickle cell haemoglobin (HbS), and mouse haemoglobin (MHb). In good agreement with the mRNA analysis, mice 29 and 54 have a large haemoglobin component that focuses at the same isoelectric point (pI) as HbS from a patient with sickle cell anaemia, whereas mouse 75 (mosaic) shows a minor component (Fig. 2d). Under these conditions we cannot distinguish mouse $\alpha\beta^s$ -dimers from human $\alpha\beta^s$ dimers.

To determine the haematological consequences of the presence of human HbS in mouse red cells, we measured the haemoglobin levels and red cell indices (Table 1). It is clear that the transgenic mice do not have anaemia and the only haematological abnormality, apart from morphological changes (see below) is a slightly lower MCV in mouse 29 (Table 1). We then performed a conventional 'sickling test'¹³. We incubated peripheral blood samples from transgenic and control animals under a sealed cover slip with the reducing agent sodium metabisulphite, and followed the changes in red cell morphology by microscopy. In mouse 29, nearly 100% of red cells sickled (Fig. 3c), like those of a human HbAS heterozygote (Fig. 3a), whereas cells from a control mouse (Fig. 3b) or mouse 54 did not sickle. It is of interest that 3–4% of the cells from the mosaic mouse 75 also sickled (data not shown); we therefore conclude that its transgenic red cells, like the red cells of mouse 29, have a sufficiently high ratio of human β^s -globin to mouse β -globin (>80%) to allow sickling. By contrast, mouse 54 still expresses sufficient levels of mouse globins to prevent sickling¹⁴.

To determine if erythrocyte sickling was occurring *in vivo*, we examined peripheral blood films. No sickle cells were found in nontransgenic animals (Fig. 3d) or in mouse 41. By contrast, sickle cells were present in mouse 29 at levels of about 1 in every 1,000 cells (Fig. 3e, f). Because we prepared the films after thorough oxygenation of the blood samples, these cells are by definition irreversibly sickled cells (ISCs)¹⁵. This is confirmed by the finding of fibrillar structures in the red cells of mouse 29, characteristic of HbS polymers¹⁶ (Fig. 4a, b). ISCs are believed to have suffered dehydration and membrane damage resulting from repeated cycles of sickling and unsickling¹⁷. Given the short half-life of mouse erythrocytes relative to that of human

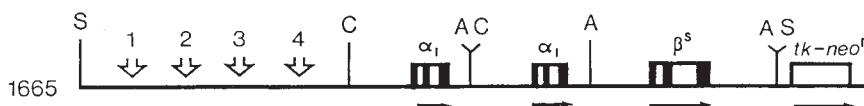


FIG. 1 Dominant control region $\alpha\beta^s$ -gene construct 1665. Plasmid vector sequences are not shown; the orientation of the genes is denoted by horizontal arrows; the vertical arrows represent the erythroid-specific DNaseI hypersensitive sites. A, Asp718 site; C, ClaI site; S, SstII site. METHODS. A 3.0-kilobase (kb) ClaI–XbaI fragment of the human α_1 -globin gene from construct 1254 (ref. 8) was cloned in the vector KS⁺ (Stratagene) and subsequently cloned between ClaI and Asp718 sites of the microlocus vector 1417^{7,12}. An NcoI–XbaI fragment of a β^s -globin allele was cloned into an Asp718–linked 4.8 kb BglII fragment of the human β^s -globin gene.

The presence of the sickle mutation was confirmed by *SauI* digestion and the Asp718 fragment cloned into the microlocus α -construct (construct 1642). The 3.0-kb ClaI–Asp718 human α_1 -globin gene fragment was ligated to ClaI linkers and cloned into the unique ClaI site of construct 1642 to give construct 1665. The final construct was digested with SstII and a 17.5-kb fragment purified for microinjection into fertilized mouse eggs. Abbreviation: tk–Neo^r, thymidine kinase gene promoter, linked to neomycin-resistance gene.

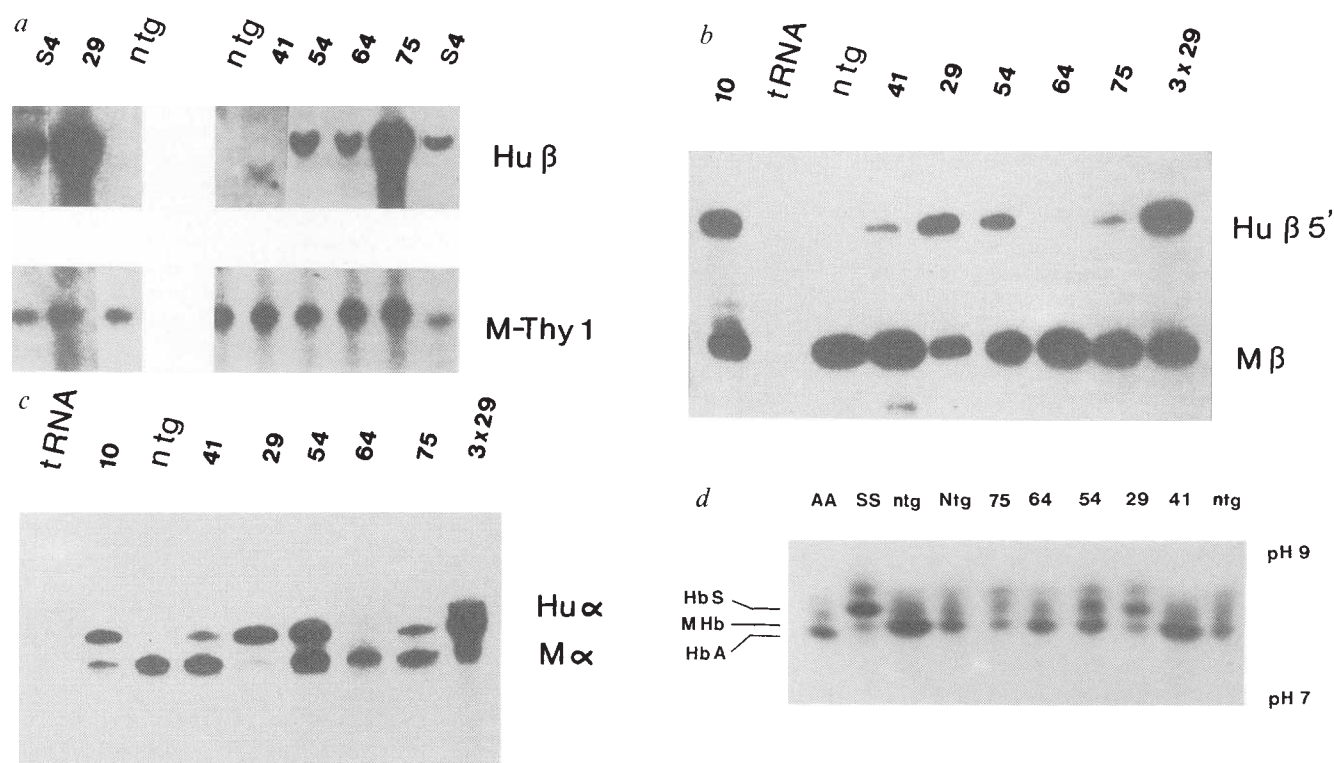


FIG. 2 Copy number, RNA and haemoglobin analysis of transgenic mice. H, human; M, mouse; ntg, nontransgenic. *a*, Tail biopsy DNAs of transgenic mice were digested with *Pst*I, fractionated by electrophoresis, transferred to nitrocellulose and hybridized to human β -globin and mouse Thy-1-specific probes. Placental DNA of a transgenic mouse (S4) with four copies of a human β -globin construct was used as a copy-number standard. *b*, Peripheral blood RNA samples (1 μ g) were analysed using a 5' human β -globin *Acc*I fragment (525 base pairs) and a mouse β -globin *Nco*I-*Hind*III probe (700 base pairs) in an S1 nuclease protection assay; specific activity ratio of the probes, 1:2 (human:mouse). Spleen RNA of a five-copy transgenic $\beta^A\alpha\theta$ transgenic mouse line (10; ref. 8) was used as positive control; nontransgenic blood RNA (ntg) and 10 μ g transfer RNA (tRNA) as negative controls. To show that the assay was performed in conditions of probe excess, 1 and

3 μ g of mouse 29 RNA were analysed (3 $\times$ 29). *c*, Peripheral blood RNA (1 μ g) was analysed with the same control RNAs as in *b*, using a 3' human α -globin *Bst*EII probe (700 base pairs) and a mouse α -globin *Bam*HI probe (300 base pairs). Specific activity ratio of the probes, 2:1 (human:mouse). *d*, Haemolysates of $\alpha\alpha\beta^S$ -transgenic mice were analysed on an LKB Multiphor IEF horizontal slab gel. Sick cell patient (SS), normal human blood (AA) and normal mouse blood (ntg) were run as markers. HbS, HbA and M Hb indicate the position of sickle cell Hb, normal human adult Hb and mouse adult Hb, respectively. The gel was photographed without staining. Gel slices were excised, the haemoglobin was eluted into water and quantitated by absorption spectroscopy at 415 nm (Table 1). METHODS. Southern blots, S1 analyses and IEF were performed as described previously⁸.

erythrocytes^{18,19}, the presence of even a few ISCs proves that repeated sickling must have occurred *in vivo*.

The fact that mouse 29, with 83% Hb^S, had a normal somatic development, no obvious manifestations of disease, and no evidence of haemolytic anaemia is interesting and constitutes a difference from the homozygous sickle cell disease. We are not sure about the reason for this discrepancy, but one explanation might be that, because of its physico-chemical properties¹⁴, 17% mouse Hb is still sufficient to protect the animal to some extent. It is remarkably reminiscent of the well-characterized human condition of heterozygosity for β^S and HPFH^{2,20} which results

in HbS levels ranging from 65 to 90%, no anaemia, normal red cell indices and mostly normal reticulocyte counts; just as mouse 29. It appears that the 17% mouse Hb does not prevent HbS polymer formation in mouse 29 (or 75, Fig. 4), but may protect it against haemolytic anaemia, just as similar levels of fetal haemoglobin protects the human β^S /HPFH heterozygotes. However, joint pains and infarctive phenomena have been reported in these subjects^{21,22}, indicating that sickling does take place in them *in vivo*, just as in mouse 29. The proportion of HbS could be increased further by crossing these transgenic mice with thalassaemic mice^{23,24}. Alternatively, transgenic lines could

TABLE 1 Globin gene expression and haematological analysis of $\alpha\alpha\beta^S$ -transgenic mice

Mouse no.	copy no. $\alpha\alpha\beta^S$ -gene	Hu β^S /M β mRNA	Hu α /Ma mRNA	HbS (% of total Hb)	Age (weeks)	Hb (g dl ⁻¹)	PCV (l l ⁻¹)	MCV (fl)	MCH (pg)	MCHC (g dl ⁻¹)	Reticulo-cytes (%)	Sickling test (%)	ISC
41	1*	0.11	0.17	0	6			58	15.7				
29	8-10	2.40	2.68	83%	15	14.4	0.49	47	13.8	29.4	ND	>95.0	+
29	(Sample I)												
29	(Sample II)	ND	ND	ND	30	13.5	0.41	45	14.6	32.7	4.4	>95.0	+
54	5	0.50	0.58	35%	4	12.5	0.46	53	14.5	27.4	8.1	0	—
64	Mosaic	0.00	0.00	0	4	14.4	0.50	58	16.9	28.8	9.0	0	—
75	Mosaic	0.17	0.24	10%	4	11.5	0.40	66	19.0	28.9	12.0	3.3	—
						13.4	0.46	53	15.9	29.8	5.3	0	—
NtC						($\pm$ 2.2)	($\pm$ 0.07)	($\pm$ 3)	($\pm$ 1.0)	($\pm$ 1.5)	($\pm$ 1.7)		

NtC, nontransgenic controls ($n=16$; 4-15 weeks; number in parentheses, s.d.); ND, not determined; Hb, haemoglobin concentration; PCV, packed cell volume; MCH, mean corpuscular haemoglobin; Hu, human; M, mouse

* Deletion in the β -globin gene (see text).

FIG. 3 Haematological analysis of transgenic mice. Sickie tests¹³ were performed on blood of a sickle trait patient (a), peripheral blood of a non-transgenic mouse (b) and peripheral blood of $\alpha\alpha\beta^3$ -transgenic mouse 29 (c). May-Grunwald Giemsa-stained blood films are shown for nontransgenic mouse blood (d) and for blood of transgenic mouse 29 (e, f).

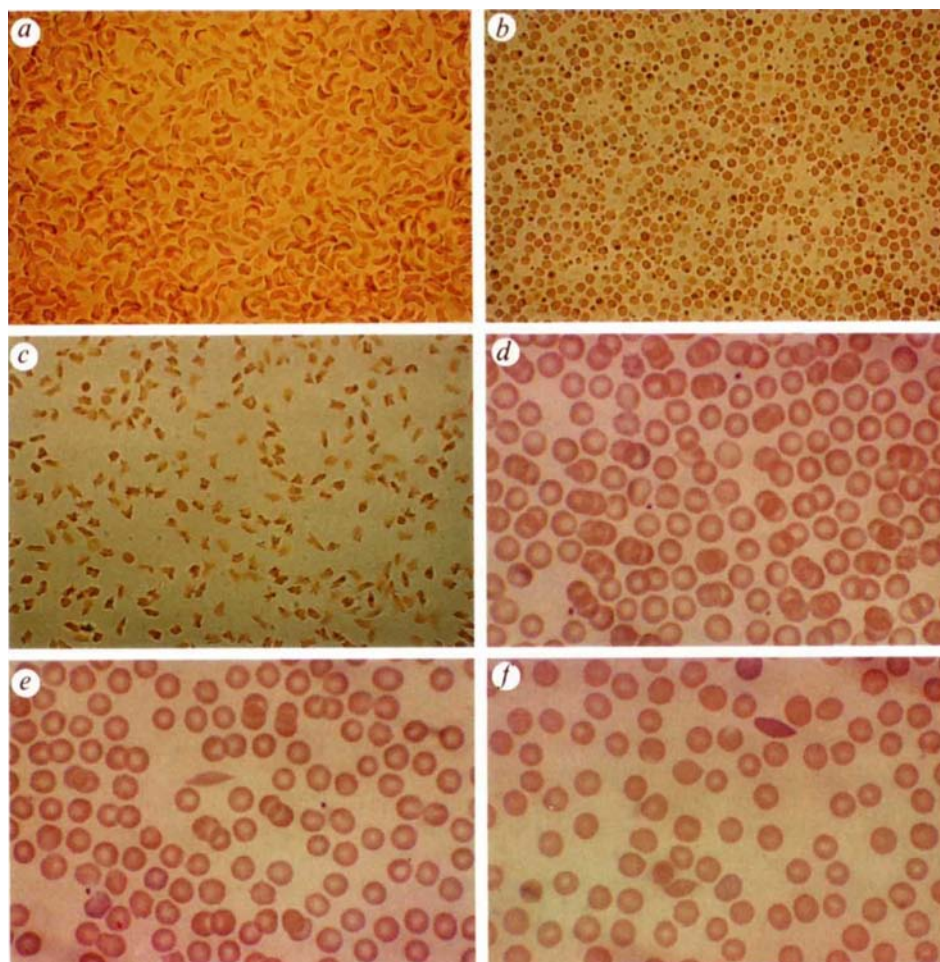
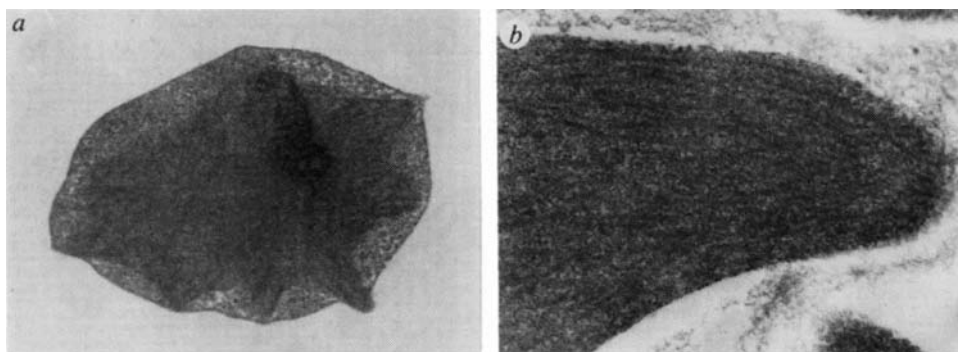


FIG. 4 a, Intracellular HbS polymers in mouse red cells. Fibres are seen in large portions of a cell which would not be classified as sickled by light microscopy (original magnification, $\times 10,000$). b, Numerous parallel polymer fibres are seen in a knob protruding from a cell which appears to be in the process of sickling (original magnification, $\times 40,000$).



be obtained with higher copy numbers. If either of these approaches will mimic not only sickling but also human sickle cell anaemia remains to be seen.

The present model should already make it possible to carry out an analysis of the factors which precipitate sickling *in vivo* and would also provide a means of screening for antisickling agents which might be clinically useful. After crossing with thalassaemic mice^{23,24} the low copy number mice should provide an excellent animal model to develop somatic gene (addition) therapy by using the β -globin DCR. $\square$

Received 18 September; accepted 1 December 1989.

- Ingram, V. A. *Nature* **178**, 792 (1956).
- Serjeant, G. R. *Sickle Cell Disease* (Oxford University Press, 1985).
- Schechter, A. N., Noguchi, C. T. & Rodgers, G. P. In *The Molecular Basis of Blood Diseases* (eds Stamatoyannopoulos, G., Nienhuis, A. W., Leder, P. & Majerus, P. W.) 179-218 (W. B. Saunders, Philadelphia, 1987).
- Luzzatto, L. & Goodfellow, P. *Nature* **337**, 17-18 (1989).
- Grosfeld, F., Blom van Assendelft, G., Greaves, D. R. & Kollias, G. *Cell* **51**, 975-985 (1987).

- Blom van Assendelft, G., Hanscombe, O., Grosfeld, F. & Greaves, D. R. *Cell* **56**, 969-977 (1989).
- Talbot, D. *et al. Nature* **338**, 352-355 (1989).
- Hanscombe, O. *et al. Genes Dev.* **3**, 1572-1581 (1989).
- Ryan, T. M. *et al. Genes Dev.* **3**, 314-323 (1989).
- Behringer, R. R. *et al. Science* **245**, 971-973 (1989).
- Ryan, T. M., Behringer, R. R., Townes, T. M., Palmiter, R. D. & Brinster, R. L. *Proc. natn. Acad. Sci. U.S.A.* **86**, 37-41 (1989).
- Collis, P., Antoniou, M. & Grosfeld, F. *EMBO J.* (in the press).
- Daland, Q. A. & Castle, W. B. *J. Lab. clin. Med.* **33**, 1082-1088 (1948).
- Rhoda, M. D. *et al. Biochim. biophys. Acta* **953**, 208-212 (1988).
- Bentley, J. F. & Milner, D. F. A. *J. clin. Invest.* **47**, 1731-1741 (1968).
- Noguchi, C. T. & Schechter, A. N. *Blood* **58**, 1057-1068.
- Padilla, F., Bromberg, P. A. & Jensen, W. N. *Blood* **41**, 653-660 (1978).
- van Ehrenstein, G. *Acta physiol. Scand.* **44**, 80-91 (1948).
- van Patten, L. M. *Blood* **13**, 789-794 (1958).
- Edington, G. M. & Lehmann, H. *Br. Med. J.* **i**, 1308-1311 (1955).
- Conley, C. L., Weatherall, D. J., Richardson, S. N., Shepard, M. K. & Charache, S. *Blood* **21**, 261-281 (1963).
- Talbot, J. F., Bird, A. C. & Sarjeant, G. R. *Br. J. Ophthalmol.* **67**, 777-778 (1983).
- Martinell, J., Whitney, J. B. III, Popp, R. A., Russell, L. B. & Anderson, W. F. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5056-5060 (1981).
- Skow, L. C. *et al. Cell* **34**, 1043-1052.

ACKNOWLEDGEMENTS. We thank Tony Magee, Olivia Hanscombe and Elaine Dzierzak for advice and reagents, Davis Swirsky for photomicrographs, and Sue Saunders for help with haemoglobin studies. P.F. is the recipient of a Cooley's Anaemia Fellowship. This work was supported by the MRC (UK).

Abscissic acid-induced elevation of guard cell cytosolic Ca^{2+} precedes stomatal closure

M. R. McAinsh*, C. Brownlee† & A. M. Hetherington*

* Institute of Environmental and Biological Sciences, Division of Biology, University of Lancaster, Lancaster LA1 4YQ, UK

† Marine Biological Association, The Laboratory, Citadel Hill, Plymouth PL1 2PB, UK

STOMATA allow the diffusion of CO_2 into the leaf for photosynthesis and the diffusion of H_2O out of the leaf during transpiration^{1,2}. This gaseous exchange is regulated by pairs of guard cells that surround each stomatal pore. During water stress the loss of water through transpiration is reduced in response to abscisic acid³, a naturally occurring plant growth regulator which is also present in certain mammals⁴, algae⁵ and fungi⁶, by the promotion of stomatal closure and inhibition of opening⁷. This involves alterations to guard cell turgor, causing the cells to shrink and thereby reducing the size of the stomatal pore. These changes are driven by cation and anion effluxes⁸. It has been proposed that an abscisic acid-dependent increase in the concentration of guard cell cytosolic free calcium triggers the intracellular machinery responsible for stomatal closure⁹ (for a review, see ref. 10), but attempts to test this hypothesis by measuring [^{45}Ca] fluxes have produced equivocal results¹¹. Using the fluorescent calcium indicator fura-2, we report that abscisic acid induces a rapid increase in guard cell cytosolic free Ca^{2+} in *Commelina communis* L., and that this increase precedes stomatal closure. These results strongly support the suggestion that Ca^{2+} is an intracellular second messenger in this response.

Figure 1 shows a typical dose response to applied abscisic acid (ABA). The minimum concentration of (+/-) *cis-trans* ABA (10^{-6} M) that induces maximum stomatal closure was used in all subsequent experiments on the effect of ABA on guard cell cytosolic free calcium ($[\text{Ca}^{2+}]_{\text{cyt}}$). To monitor real time changes in $[\text{Ca}^{2+}]_{\text{cyt}}$ in single stomatal guard cells, we used the fluorescent calcium indicator fura-2¹². Fluorescent indicators such as fura-2 and indo-1 have already been used successfully to monitor free calcium concentrations in root hairs¹³ and aleurone layer protoplasts^{14,15}. Preliminary experiments established that stomatal guard cells in common with other plant cells (but unlike many animal cells) would not take up the esterified form of this molecule. Additionally, the failure of the low-pH loading method, successfully used with plant protoplasts¹⁵, necessitated that the dye be micro-injected into the cytoplasm using iontophoresis. Figure 2a shows a bright-field image of a stomatal complex, the right-hand guard cell of which has been loaded with fura-2. In Fig. 2b the same complex is

shown under fluorescence and it is apparent that the fura-2 has been loaded into the cytoplasm of the right-hand cell. Guard cells with fura-2 loaded into the cytoplasm remained viable throughout the course of the experiments and continued to respond to ABA by reducing their turgor. Furthermore, such cells retained dye in the cytoplasm with no obvious accumulation in the vacuole. By contrast, guard cells in which fura-2 had been loaded into the vacuole immediately lost turgor and the distribution of the dye appeared to be much more diffuse when viewed under fluorescence. These factors meant that we were able to reject any cells in which the dye was not loaded into the cytoplasm.

When guard cells previously loaded with fura-2 were perfused with 10^{-6} M ABA, there was an immediate increase in $[\text{Ca}^{2+}]_{\text{cyt}}$ (Fig. 3b); this increase preceded any detectable alteration in the aperture of the stomatal pore by at least 6 min (Fig. 3c). It was important that the $[\text{Ca}^{2+}]_{\text{cyt}}$ in untreated guard cells remained constant (Fig. 3a) and there was no alteration in the aperture of the stomatal pore. Using an *in vitro* calibration curve (see legend to Fig. 3), we estimate that resting $[\text{Ca}^{2+}]_{\text{cyt}}$ lies between 70–250 nM, with a mean of 115 ± 26 nM ($n = 10$). Eight out of ten guard cells responded to 10^{-6} M ABA with an increase in $[\text{Ca}^{2+}]_{\text{cyt}}$ (there was no increase or decrease in $[\text{Ca}^{2+}]_{\text{cyt}}$ in the remaining two cells). These responses represent an increase over resting $[\text{Ca}^{2+}]_{\text{cyt}}$ of between 2- and 10-fold. For example, in the response shown in Fig. 3, the $[\text{Ca}^{2+}]_{\text{cyt}}$ rose from a resting $[\text{Ca}^{2+}]_{\text{cyt}}$ of 70 nM to 600 nM, with peaks up to 1 μM , about 10 min after the addition of ABA. Variations in $[\text{Ca}^{2+}]_{\text{cyt}}$ which resemble oscillations were also apparent after ABA treatment (Fig. 3b). Interestingly, these are more similar in form to those in mammalian cells¹⁶ than those induced by auxin in maize epidermal cells¹⁷, but further investigation is necessary to characterize this phenomenon in stomatal guard cells.

The absolute values of 360 nm fluorescence (fura-2 excitation isosbestic point) for fura-2-loaded guard cells (Fig. 3d) showed no marked change following the addition of ABA. There was a slight decrease in fluorescence with time, which we attribute to either a gradual leakage of dye from the cell and/or dye bleaching, both of which are compensated for by the dual wavelength ratio measurements¹². This confirms that the observed increases in $[\text{Ca}^{2+}]_{\text{cyt}}$ are not an artefact due to modification to the optical properties of the cell brought about by ABA or guard cell movement during stomatal closure.

MacRobbie observed a transient stimulation of ion effluxes in response to ABA (refs 11, 18) and has suggested that this could provide a mechanism by which increases in $[\text{Ca}^{2+}]_{\text{cyt}}$ might initiate stomatal closure. The similarity between the time course of these efflux transients^{11,18} and the ABA-induced increase in $[\text{Ca}^{2+}]_{\text{cyt}}$ that we report here, strongly suggests that these two events are interrelated. Recent patch clamping studies^{19,20} also help to explain the mechanism by which an increase in $[\text{Ca}^{2+}]_{\text{cyt}}$ might lead to a reduction in aperture of

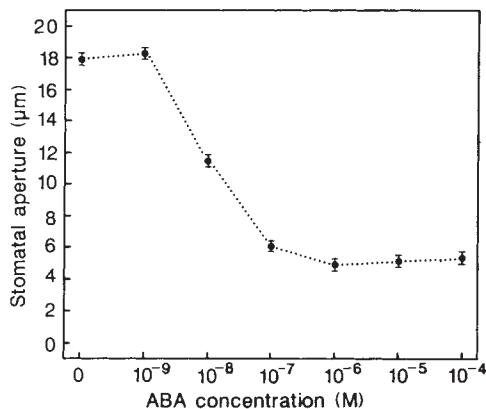


FIG. 1 Response of abaxial stomata of *C. communis* L. to a range of concentrations of abscisic acid (10^{-9} – 10^{-4} M). Means of 90 measurements $\pm$ s.e. Maximum response was observed between 10^{-7} and 10^{-6} M ABA.

METHODS. Seedlings of *C. communis* L. were grown from seed⁹. Epidermis was peeled carefully from the abaxial surface of the youngest fully expanded leaf²², floated on 10^{-2} M MES (2-[N-morpholino]ethane sulphonic acid; Fluka), pH 6.15, and cut into 5 mm lengths. To promote stomatal opening, pieces of epidermis were incubated for 3 h at $25 \pm 1^\circ\text{C}$ under a photon flux density of $200 \mu\text{mol m}^{-2} \text{s}^{-1}$ in 10^{-2} M MES, 5×10^{-2} M KCl, pH 6.15, aerated with CO_2 -free air⁹, after which (+/-) *cis-trans* ABA (Sigma) was added to the medium. Following incubation for 1 h in ABA, the epidermal strips were examined under the microscope to determine the aperture of the stomatal pores.

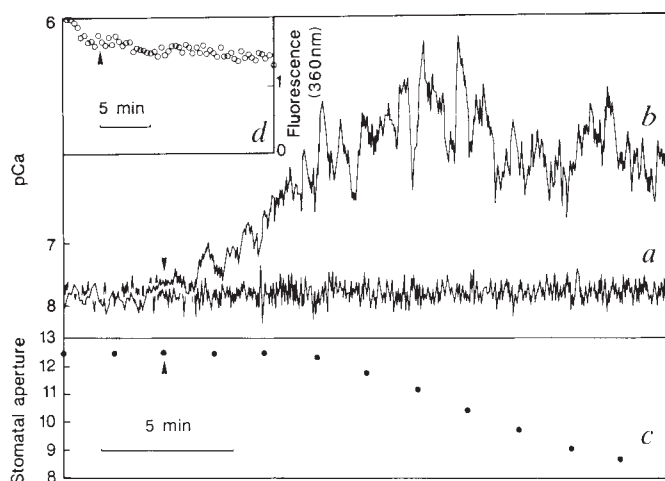


FIG. 2 *a*, Bright-field image of a stomatal complex from *C. communis* L. The diameter of the pore is 12 μm (see legend to Fig. 1 for methods). After loading with fura-2, the right-hand guard cell retained a constant aperture for 45 min and dye remained in the cytoplasm. *b*, Fluorescence image of the same fura-2-loaded guard cell as in *a*. Excitation was at 360 nm. Emission was monitored at 510 nm (see legend to Fig. 3).

METHODS. Epidermal strips of >2 cm in length were incubated under conditions promoting stomatal opening for 2–3 h (see legend to Fig. 1). Strips on which the stomata were open (12–15 μm) were mounted, cuticle down, on a glass coverslip and secured by smaller coverslips. The preparation was seated in a purpose-built perfusion cell on the microscope stage and the strip perfused continuously with CO_2 -free 10^{-2} M MES, 5×10^{-2} M KCl, pH 6.15, at 25 $^{\circ}\text{C}$. This treatment resulted in no alteration to the stomatal aperture. Fura-2 was microinjected iontophoretically. Guard cells were impaled with filamented glass microelectrodes (<0.25 μm tip diameter) containing 10^{-3} M fura-2 pentapotassium salt (Calbiochem) in their tips²³. The electrodes were connected with silver/silver chloride wire to a microelectrode amplifier and stimulator. Current pulses (1.0 nA negative pulses, 2 Hz, 200 ms duration) were given for <30 s to load cells, giving a final estimated concentration of fura-2 in the cytoplasm of $\sim 10^{-4}$ M (ref. 24). Following microinjection, the electrode was removed for fluorescence measurements. Fluorescence was observed with a $\times 40$ water-immersion objective (Zeiss).

FIG. 3 Changes in guard cell $[\text{Ca}^{2+}]_{\text{cyt}}$ with time. *a*, Cells perfused with CO_2 -free 10^{-2} M MES, 5×10^{-2} M KCl, pH 6.15, at 25 $^{\circ}\text{C}$. *b*, Before and after addition of 10^{-6} M ABA (arrow). *c*, Changes in stomatal aperture with time, before and after addition of 10^{-6} M ABA (arrow). *d*, Changes in 360 nm fluorescence (fura-2 excitation isosbestic point) with time, before and after the addition of 10^{-6} M ABA (arrow).

METHODS. Stomatal guard cells of *C. Communis* L. were loaded with fura-2 (see legend to Fig. 2). Ca^{2+} -dependent fura-2 fluorescence was monitored with dual wavelength fluorescence microscopy^{23,25–27}. Fura-2-loaded cells were observed with a modified M2 micromanipulation microscope (Microinstruments, UK), using a $\times 40$ water-immersion objective (Zeiss). Excitation, using 360 and 385 nm light alternately, was provided from a 150 watt xenon lamp source by a fluid light guide (Microinstruments) in combination with a rotating filter holder. Fluorescence emission at 510 nm was monitored with a photomultiplier tube (PM9924B; EMI), the output of which was synchronized with the excitation source by a BBC microcomputer. Each fluorescence measurement was an average of 50 individual readings. Guard cell autofluorescence was measured at each wavelength before loading with fura-2. Autofluorescence subtraction and 360/385 nm ratio calculations were performed on-line during the experiment to give one ratio reading every 2 s. The aperture of the stomatal pore was determined *in situ* every 50 ratios (~ 100 s) using a calibrated eye piece micrometer. Values of pCa are based on *in vitro* calibration using drops of Ca^{2+} -calibration buffers containing 2×10^{-5} M fura-2 (ref. 28). We found that *in vivo* calibration with ionomycin^{23,25,26} was not possible. No changes in $[\text{Ca}^{2+}]_{\text{cyt}}$ were observed in guard cells from epidermal strips incubated in solutions of ionomycin contain-



ing zero (10^{-3} M EGTA) or 10^{-2} M CaCl_2 , pH 8.0. Furthermore, the combination of high $[\text{Ca}^{2+}]$ and ionomycin failed to induce stomatal closure and did not influence cytoplasmic streaming, which is known to be strongly Ca^{2+} -dependent²⁹.

the stomatal pore. An increase in $[\text{Ca}^{2+}]_{\text{cyt}}$ would be expected to block inward rectifying K^+ channels and hence inhibit stomatal opening, whereas closure might result from the activation of a calcium-stimulated chloride channel. The resulting chloride efflux could then depolarize the guard cell plasma membrane sufficiently to activate the outwardly directed K^+ channel^{19,20}.

In summary, we have reported that ABA induces an increase

in guard cell $[\text{Ca}^{2+}]_{\text{cyt}}$ which precedes stomatal closure. Our results provide an important link in the guard cell signal transduction chain. It will be of interest to determine whether opening stimuli, such as light and auxin, induce alterations to guard cell $[\text{Ca}^{2+}]_{\text{cyt}}$. Experiments using calcium-sensitive microelectrodes have revealed that light causes a decrease in $[\text{Ca}^{2+}]_{\text{cyt}}$ in *Nitellopsis*¹⁹, and auxin causes an increase in $[\text{Ca}^{2+}]_{\text{cyt}}$ in maize¹⁷. Future

work will attempt to establish the site(s) of calcium entry into the cytosol, including the involvement of intracellular stores, and also to identify other components in the guard cell signal transduction chain. □

Received 8 September; accepted 23 November 1989.

1. Raschke, K. in *Encyclopedia of Plant Physiology* New Ser. Vol. 7 (eds Haupt, W. & Feinleib, M. E.) 381–441 (Springer, Berlin, 1979).
2. Zeiger, E. A. *Rev. Plant. Physiol.* **34**, 441–475 (1983).
3. Davies, W. S. & Mansfield, T. A. in *Abscissic Acid* (ed. Addicott, F. T.) 237–268 (Praeger, New York, 1983).
4. Le Page-Degivry, M.-Th., Bidard, J.-N., Rouvier, E., Boulard, C. & Lazdunski, M. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1155–1158 (1986).
5. Tietz, A., Rutkowski, U., Kohler, R. & Kasprisk, W. *Biochem. Physiol. Pflanz.* **184**, 259–266 (1989).
6. Assante, G., Merlini, L. & Nasini, G. *Experientia* **33**, 1556 (1977).
7. Mittelheuser, C. G. & van Steveninck, R. F. M. *Nature* **221**, 281–282 (1969).
8. MacRobbie, E. A. C. in *Stomatal Function* (eds Zeiger, E., Farquhar, G. D. & Cowan, I. R.) 125–162 (Stanford University Press, California, 1987).
9. De Silva, D. L. R., Hetherington, A. M. & Mansfield, T. A. *New Phytol.* **100**, 473–482 (1985).
10. Mansfield, T. A., Hetherington, A. M. & Atkinson, C. J. A. *Rev. Pl. Physiol.* (in the press).
11. MacRobbie, E. A. C. *Planta* **178**, 231–241 (1989).
12. Tsien, R. Y. A. *Rev. Neurosci.* **12**, 227–254 (1989).
13. Clarkson, D. T., Brownlee, C. & Ayling, S. M. *J. Cell. Sci.* **91**, 71–80 (1988).
14. Bush, D. S. & Jones, R. L. *Cell Calcium* **8**, 455–472 (1987).
15. Bush, D. S. & Jones, R. L. *Eur. J. Cell Biol.* **46**, 446–469 (1988).
16. Berridge, M. S., Cobbold, P. H. & Cuthbertson, K. S. R. *Phil. Trans. R. Soc. B* **320**, 325–343 (1988).
17. Felle, H. *Planta* **174**, 495–499 (1988).
18. MacRobbie, E. A. C. *Bot. Acta* **101**, 140–148 (1988).
19. Schroeder, J. I. & Hedrich, R. *Trends Biochem. Sci.* **14**, 187–192 (1989).
20. Schroeder, J. I. & Hagiwara, S. *Nature* **338**, 427–430 (1989).
21. Miller, A. J. & Sanders, D. *Nature* **326**, 397–400 (1987).
22. Weyers, J. D. B. & Travis, A. J. *J. exp. Bot.* **32**, 837–850 (1981).
23. Brownlee, C. & Pulsford, A. L. *J. Cell Sci.* **91**, 249–256 (1988).
24. Purves, R. D. *Microelectrode Methods for Intracellular Recording and Ionophoresis* (Academic, New York, 1981).
25. Williams, D. A., Fogarty, K. E., Tsien, R. Y. & Fay, F. S. *Nature* **318**, 558–561 (1985).
26. Poenie, M., Alderton, S., Steinhardt, R. & Tsien, R. *Science* **233**, 886–889 (1986).
27. Brownlee, C. *Biol. Bull.* **176**(s), 14–17 (1989).
28. Tsien, R. Y. & Rink, T. J. *Biochim. biophys. Acta* **599**, 623–638 (1980).
29. Williamson, R. E. & Ashley, C. C. *Nature* **296**, 647–651 (1982).

ACKNOWLEDGEMENTS. We thank the Agricultural and Food Research Council, The Gatsby Charitable Foundation, The University of Lancaster Committee for Research and The Marine Biological Association of the UK for providing financial assistance.

Role for the nitrogenase MoFe protein α -subunit in FeMo-cofactor binding and catalysis

Deborah J. Scott*, Harold D. May*, William E. Newton*†, Kevin E. Brigle‡ & Dennis R. Dean†‡

* Western Regional Research Center, USDA-ARS, Albany, California 94710, USA and Department of Agronomy and Range Science, University of California at Davis, Davis, California 95616, USA

‡ Department of Anaerobic Microbiology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, USA

TWO components constitute Mo-dependent nitrogenase (EC 1.18.6.1)—the Fe protein (a homodimer encoded by *nifH*) and the MoFe protein (an $\alpha_2\beta_2$ tetramer encoded by *nifDK*). The MoFe protein provides the substrate-binding site^{1–3} and probably contains six prosthetic groups of two types—four Fe-S centres and two Fe- and Mo-containing cofactors^{4,5}. To determine the distribution and catalytic function of these metalloclusters, we^{6,7} and others⁸ are attempting to change the catalytic and spectroscopic features of nitrogenase by substituting specific amino acids targeted as potential metallocluster ligands, particularly those to the FeMo-cofactor, which is responsible for the biologically unique

electron paramagnetic resonance signal ($S = \frac{3}{2}$) of nitrogenase^{9,10}, and is believed to be the N_2 -reducing site¹¹. Here we describe mutant strains of *Azotobacter vinelandii* that have single amino-acid substitutions within the MoFe protein α -subunit. These substitutions alter both substrate-reduction properties and the unique electron paramagnetic resonance signal, indicating that the FeMo-cofactor is associated with both the α -subunit and the substrate-reducing site.

Nitrogenase MoFe proteins all have a very high level of primary sequence identity. To target appropriate residues for substitution, we correlated the chemical nature of the denaturing solvents that are effective in extruding each type of metallocluster with comparisons of interspecies and intersubunit amino-acid sequences. Residues targeted for substitution include certain strictly conserved cysteine and histidine residues⁷ that could respectively serve as thiol or N-donor ligands to the FeMo-cofactor. Histidine residues are known to ligand protein-bound Fe-S clusters¹². Also, metallocluster extrusion requirements^{9,13}, electron spin echo¹⁴ and Fourier transform infrared analyses¹⁵ indicate one or more deprotonated nitrogen ligands to the FeMo-cofactor.

Primary sequences of those *nif*-specific gene products required for the biosynthesis of the FeMo-cofactor have also been examined to gain an insight into the possible nature of the binding domains for the FeMo-cofactor that are located within the MoFe protein. For example, we have suggested that the products of two such FeMo-cofactor biosynthetic genes, *nifE* and *nifN*, form a scaffold on which the FeMo-cofactor is assembled¹⁶. Indeed, the *nifE* and *nifN* gene products have considerable identity with the MoFe protein α - and β -subunits, respectively^{16–21}. Also, the *nifEN* gene products form a tetrameric complex²² analogous to the MoFe protein. Because the FeMo-cofactor must escape from the NifEN biosynthetic complex during maturation of the MoFe protein, the respective FeMo-cofactor-binding sites of the MoFe protein and NifEN are likely to be structurally similar but functionally inequivalent. We, therefore, substituted histidine 195 of the MoFe protein α -subunit with the corresponding residue of the *nifE* gene product, asparagine (Fig. 1). In this way, we expected to alter the functional character of a potential FeMo-cofactor environment without severely altering the global polypeptide structure. We incorporated amino-acid substitutions using site-directed mutagenesis and a gene replacement procedure⁶.

The resulting *A. vinelandii* mutant strain (DJ178) had a Nif[−] phenotype. We compared the catalytic activity of its altered nitrogenase with that of native nitrogenase from the isogenic wild-type strain. In an atmosphere of 10% acetylene, wild-type crude extracts catalysed the reduction of acetylene to ethylene with no trace of ethane production. By contrast, the nitrogenase in crude extracts of DJ178, under identical conditions, produced ethane as well as ethylene, with ethane accounting for 35% of the electrons going to these products (Table 1). In addition,



FIG. 1 Comparison of the amino-acid sequences of the *A. vinelandii* MoFe protein α -subunit³⁰ and the deduced *nifE* gene product sequence¹⁷ in the region relative to the His 195 $\rightarrow$ Asn and Gln 191 $\rightarrow$ Lys substitutions in the α -subunit. Numbers refer to the residue position within the respective polypeptides. The initiating methionine is residue 1. Asterisks above the MoFe protein α -subunit sequence indicate residues that are conserved among all known MoFe protein α -subunit sequences. Asterisks below the *nifE* gene product sequence indicate residues that are conserved among all known *nifE* gene product sequences. Amino acids conserved between the *A. vinelandii* MoFe protein α -subunit sequence and the *nifE* gene product sequence are boxed. ↓, Gln and His residues substituted by the corresponding *nifE* gene product residue in the study described here.

† To whom correspondence should be addressed at Department of Anaerobic Microbiology, Virginia Polytechnic Institute & State University, Blacksburg, Virginia 24061, USA

TABLE 1 Specific activities and construction of strains

Strain	Mutation*	Oligonucleotide†	Growth rate‡	MoFe protein§		Fe protein
				C ₂ H ₆	C ₂ H ₄	C ₂ H ₄
Wild type	—	—	2.5	0.0	40.6	18.7
DJ178	α -His 195 → Asn	CAGTCCCTGGCAACCACATC	NG	1.0	3.7	26.3
DJ357	$\Delta nifDK::Km^r$	—	NG	0.0	0.0	48.5
DJ255	α -Gln 191 → Lys	CGCGGCGTTTCCAAGTCCCTG	NG	0.1	1.4	39.5

* The MoFe protein α -subunit is indicated by α -, followed by the amino-acid residue substitution.

† Used to generate the codon change. The replacement base is underlined (see ref. 6).

‡ Growth rate is doubling time (h) when cultured in a minimal medium with no added fixed nitrogen source; NG, no growth.

§ Cells were derepressed and extracts prepared as previously described¹⁶, except cells were ruptured by sonication. Specific activity units are nmol product formed per min per mg total protein in the presence of saturating purified²⁸ *A. vinelandii* Fe protein.

|| As above (§), except that saturating purified²⁸ *A. vinelandii* MoFe protein was present.

¶ Strain DJ357 was constructed by transforming the phenotype of DJ178 to Km^r using the plasmid pDB253. This plasmid has as its origin plasmid pDB6 (ref. 30), which carries the *nif* structural gene cluster from *A. vinelandii*. The plasmid pDB253 has a pair of *Pst*I restriction enzyme fragments deleted and replaced with a Km^r gene cartridge. The deletion located within the DJ357 chromosome extends 1,033 base pairs and removes the $-CO_2H$ -encoding end of *nifD* and the $-NH_2$ -encoding end of *nifK*.

whole cells of DJ178 exhibited an electron paramagnetic resonance (EPR) spectrum, which was changed in both lineshape and g value when compared with that of wild-type cells (Fig. 2). The intensity of the DJ178 EPR signal was also considerably diminished.

To examine further whether this region of the α -subunit serves as an FeMo-cofactor-binding domain, we substituted lysine for glutamine 191. Again, this choice was guided by the sequence of the *nifE* gene product (Fig. 1). This mutant strain, DJ255, also had a Nif^- phenotype. Crude extracts from DJ255 reduced acetylene to ethane but at a relatively lower level (12.5% of electron flow; Table 1). The whole-cell EPR spectrum of nitrogenase of this mutant was of comparable intensity to that of wild type, but was again changed in both lineshape and g value. But these changes were different to those observed for DJ178 (Fig. 2). Two-dimensional gel electrophoretic analyses of wild-type, DJ178 and DJ255 crude extracts demonstrated the predicted shift in isoelectric point (pI) of the altered α -subunits and also revealed that no new polypeptides accumulated in the mutant strains (data not shown).

Although ethane formation from acetylene is not detected with native Mo-dependent nitrogenase, it is used as a test for V-nitrogenases *in vivo*²³. The changed catalytic and spectroscopic properties observed for DJ178 and DJ255 cannot, however, be explained by the presence of either an alternative nitrogenase or a hybrid between the Mo-dependent system and an alternative nitrogenase because: (1) nitrogenase-derepression experiments were performed under conditions which completely repress expression of alternative nitrogenase structural components²⁴; (2) polypeptides corresponding to alternative nitrogenase components do not accumulate in DJ178 or DJ255 when derepressed for nitrogenase synthesis; (3) a strain derived from DJ178, having inactivated *nifDK* genes, produces neither ethylene nor ethane (Table 1); (4) the catalytic activities and EPR spectra of V-nitrogenase^{25,26}, and wild-type, DJ178 and DJ255 nitrogenases are all different; and (5) certain other mutant strains that we have constructed, for example, with serine replacing cysteine 154 in the α -subunit, exhibit no catalytic activity and no EPR signal⁶.

Our interpretation of these observations is that the substitutions in DJ178 and DJ255 cause a replacement, displacement, or distortion of ligands to the FeMo-cofactor from the α -subunit polypeptide. Although these experiments cannot definitely pinpoint glutamine 191 or histidine 195 as direct ligands to the FeMo-cofactor, the catalytic and spectral changes caused by these substitutions indicate that the FeMo-cofactor must be associated with the α -subunit. These results, however, do not preclude the possibility that the FeMo-cofactor is also bound by the β -subunit or by other regions within the α -subunit. In fact, the region surrounding cysteine 275 within the α -subunit

could also provide an FeMo-cofactor-binding domain^{7,8,27}. The observation that amino-acid substitutions, which result in changed catalytic activities, also affect the EPR spectrum of the protein-bound FeMo-cofactor strongly supports the proposal that the FeMo-cofactor is located at or near the substrate-reduction site¹¹. Our results also indicate that the EPR spectrum and the catalytic activity of nitrogenase are influenced not only by the structure and/or composition of the FeMo-cofactor,

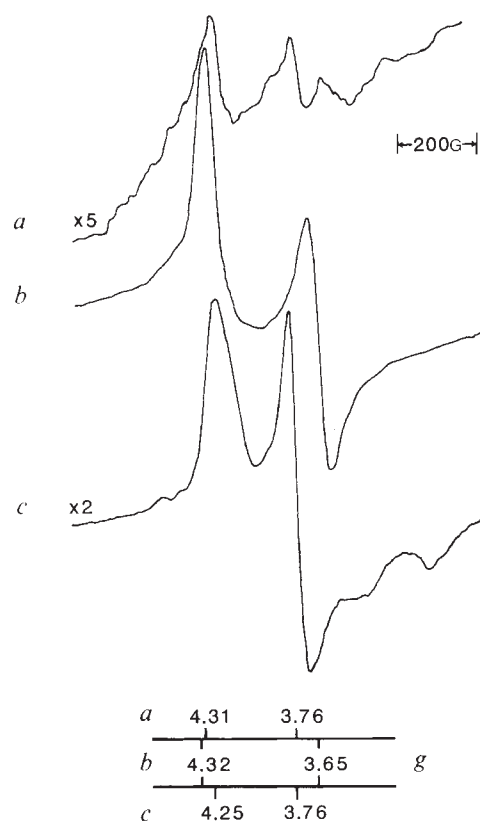


FIG. 2 EPR spectra in the $g=4$ region of a, DJ178, b, wild type and c, DJ255 whole cells. Apparent g values and relative receiver gains are indicated. METHODS. Nitrogenase-derepressed whole cells of *A. vinelandii* wild-type, DJ178 and DJ255 strains were separately treated in 50 mM Tris buffer (pH 8) with 5 mM sodium dithionite and 0.1 mM methyl viologen³¹ under argon for 5 min, followed by centrifugation at 600g into degassed, serum-stoppered EPR tubes and freezing in liquid nitrogen. Spectra were recorded at 9.23 GHz and at 50 mW on a Varian Associates Series E-Line Spectrometer at 12 K maintained by liquid helium boil-off.

but also by its polypeptide environment. Thus, an active role for the α -subunit polypeptide in nitrogenase catalysis is indicated. □

Received 8 August; accepted 14 November 1989.

- Orme-Johnson, W. H. et al. *Proc. natn. Acad. Sci. U.S.A.* **69**, 3142–3145 (1972).
- Mortenson, L. E., Zumft, W. G. & Palmer, G. *Biochim. biophys. Acta* **292**, 422–435 (1973).
- Smith, B. E., Lowe, D. J. & Bray, R. C. *Biochem. J.* **135**, 331–341 (1973).
- Smith, B. E. & Lang, G. *Biochem. J.* **137**, 169–180 (1974).
- Zimmerman, R. et al. *Biochim. biophys. Acta* **537**, 185–207 (1978).
- Brigle, K. E. et al. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7066–7069 (1987).
- Dean, D. R., Brigle, K. E., May, H. D. & Newton, W. E. in *Nitrogen Fixation: Hundred Years After* (eds Bothe, H., de Bruijn, F. J. & Newton, W. E.) 107–113 (Fischer, Stuttgart, 1988).
- Kent, H. M., Ionnidis, I., Gormal, C., Smith, B. E. & Buck, M. *Biochem. J.* **264**, 257–264 (1989).
- Shah, V. K. & Brill, W. J. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3249–3253 (1977).
- Rawlings, J. et al. *J. biol. Chem.* **253**, 1001–1004 (1978).
- Hawkes, T. R., McLean, P. A. & Smith, B. E. *Biochem. J.* **217**, 317–321 (1984).
- Telser, J. et al. *FEBS Lett.* **214**, 117–121 (1987).
- Yang, S.-S. et al. *J. biol. Chem.* **257**, 8042–8048 (1982).
- Thomann, H. et al. *J. Am. chem. Soc.* **109**, 7913–7914 (1987).
- Walters, M. A., Chapman, S. K. & Orme-Johnson, W. H. *Polyhedron* **5**, 561–565 (1986).

- Brigle, K. E., Weiss, M. C., Newton, W. E. & Dean, D. R. *J. Bact.* **169**, 1547–1553 (1987).
- Brigle, K. E. & Dean, D. R. *Nucleic Acids Res.* **16**, 5214 (1988).
- Setterquist, R. et al. *Nucleic Acids Res.* **16**, 5215 (1988).
- Arnold, W., Rump, A., Klipp, W., Priefer, U. B. & Puhler, A. *J. molec. Biol.* **230**, 715–738 (1988).
- Moreno-Vivian, C., Schmel, M., Masepohl, B., Arnold, W. & Klipp, W. *Molec. gen. Genet.* **216**, 353–363 (1989).
- Wang, S.-Z., Chen, J.-S. & Johnson, J. L. *Nucleic Acids Res.* **17**, 3299 (1989).
- Paustian, T. D., Shah, V. K. & Roberts, G. P. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6082–6086 (1989).
- Dilworth, M. J., Eady, R. R., Robson, R. L. & Miller, R. W. *Nature* **327**, 167–168 (1987).
- Jacobson, M. R., Premakumar, R. & Bishop, P. E. *J. Bact.* **167**, 480–486 (1986).
- Eady, R. R., Robson, R. L., Richardson, T. H., Miller, R. W. & Hawkins, M. *Biochem. J.* **244**, 197–207 (1987).
- Hales, G. J., Case, E. E., Morningstar, J. E., Dzeda, M. F. & Mauterer, L. A. *Biochemistry* **25**, 7251–7255 (1986).
- Brigle, K. E., Newton, W. E. & Dean, D. R. *Gene* **37**, 37–44 (1985).
- Burgess, B. K., Jacobs, D. J. & Stiefel, E. I. *Biochim. biophys. Acta* **614**, 196–209 (1980).
- Gornall, A. G., Bardawell, C. J. & David, M. M. *J. biol. Chem.* **177**, 751–766 (1948).
- Jacobson, M. R. et al. *J. Bact.* **171**, 1017–1027 (1989).
- Howard, K. S., Hales, B. J. & Socolofsky, M. D. *Biochim. biophys. Acta* **812**, 575–585 (1985).

ACKNOWLEDGEMENTS. We thank W. R. Dunham for assistance with EPR spectroscopy and Nancy Laird, Robert Setterquist and Valerie Cash for technical assistance. This work was supported by the NSF.

DNA trapping electrophoresis

Levy Ulanovsky, Guy Drouin & Walter Gilbert

Department of Cellular and Developmental Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA

ATTEMPTS to improve the size separation of single-stranded DNA in polyacrylamide gels by field-inversion gel electrophoresis (FIGE) have met with limited success¹. Here we show that attaching a neutral globular protein, streptavidin, to one end of a single-stranded DNA molecule profoundly alters the DNA mobility pattern and increases the band separation manifold within a size range controlled by voltage and pulse cycle. In constant field, short modified fragments are only slightly retarded but long molecules are retarded dramatically above a 'threshold size' of 0.6 kilobases at 60 V per cm. At this voltage, molecules above a 1.2-kilobase 'cut-off' do not enter the gel. Both the threshold and the cut-off sizes decrease as the voltage increases. In FIGE, the longer the

reverse pulse, the larger the modified fragments that enter the gel. We interpret these results as the trapping by the gel matrix of the protein attached to the DNA. The probability of release then depends on the balance between the electric field and thermal motion: the larger the DNA and the higher the voltage, the harder it is to release.

In conventional gel electrophoresis, the mobility of DNA decreases as the size of the molecule increases until a 'limiting mobility', independent of size, is reached: above that size electrophoresis offers no resolution. Figure 1 shows the electrophoresis of single-stranded DNA fragments with and without an attached streptavidin molecule. The 123-base ladders in lanes B and C show that the protein-DNA complex displays no limiting mobility, unlike naked DNA; instead, the mobility difference between adjacent bands of the ladder increases until the complex no longer enters the gel. Figure 2 plots the mobilities of the modified and naked DNA versus DNA size for different field strengths. Above a certain threshold size, which coincides with the beginning of the transition to the limiting mobility for

FIG. 1 Migration of streptavidin-labelled versus naked single-stranded DNA on *a*, 8% and on *b*, 6% polyacrylamide gels at constant voltages of 60 V per cm and 100 V per cm respectively. Lane A, SV40/*Hind*III with streptavidin; lane B, 123-base ladder (BRL) with streptavidin; lane C, naked 123-base ladder; lane D, naked SV40/*Hind*III. The numbers on the left/right indicate the sizes (bases) of the streptavidinated/naked DNA fragments. The arrows indicate the markers other than the 123-base ladder. Recessive 3' ends were filled in enzymatically with Sequenase with an inner nucleotide biotinylated, and one of the following nucleotides radioactively labelled with ³²P. Labelled samples were incubated with streptavidin (BRL) in molar excess over both incorporated and free biotinylated dNTP. The DNA was denatured by heating at 65 °C for 5 min in 80% formamide and electrophoresed on 6% or 8% polyacrylamide gels, 1:19 bisacrylamide in 7M urea. Gels were 13.5 cm long (slot-to-bottom), 0.5 mm thick, and were water-cooled (BioRad Protean-II Slab cell). In *a*, the temperature was 15 °C during the 8-h run. In *b*, it rose from 14 °C to 20 °C during the 60-min run.

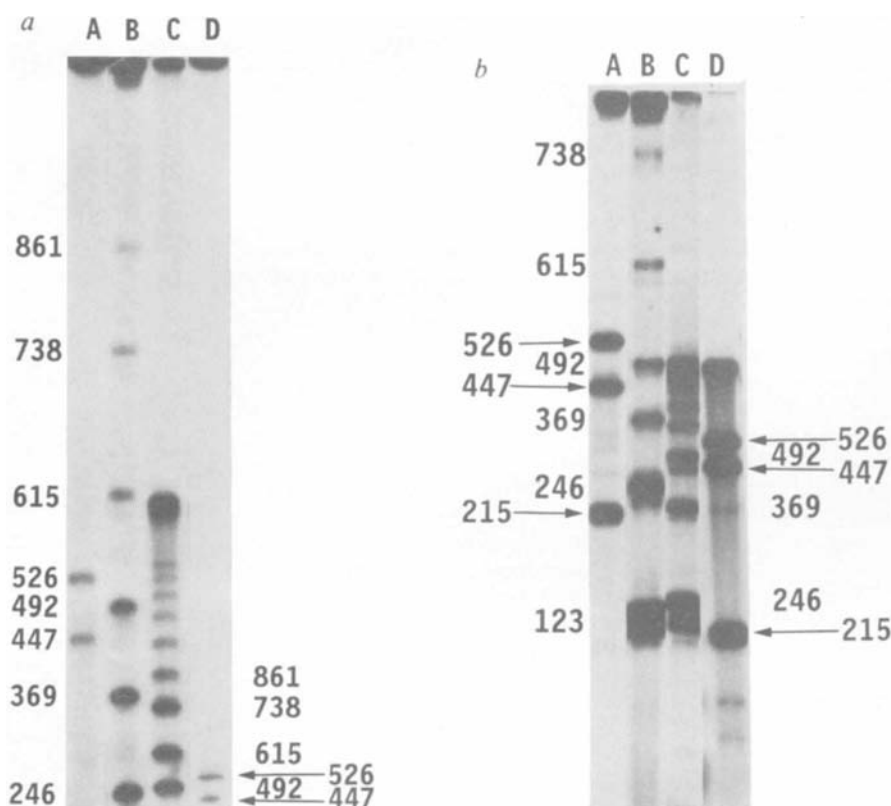
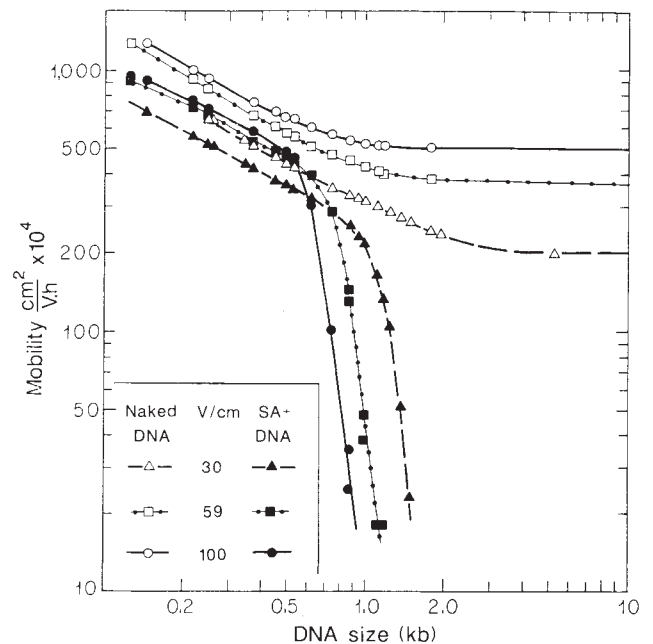


FIG. 2 Migration velocity per unit electric field as a function of DNA length for streptavidin (SA)-labelled and naked DNA on 6% polyacrylamide gels in constant fields. The voltage for the different curves is shown in the insert. Experimental procedures were as described in the legend to Fig. 1.



the naked DNA, the retardation induced by streptavidin grows rapidly. Increased voltage shifts this effect to smaller DNA sizes.

The electrically neutral streptavidin (relative molecular mass 48,000), being bulkier than the DNA chain and thus trailing behind it, might be trapped occasionally by the gel matrix. The probability of its release would then depend on the balance between the electric field, which keeps the molecule 'hooked up' in the trap by pulling it forwards, and the thermal motion, which tends to lift the molecule 'off the hook'. The force exerted on the molecule by the electric field is proportional to the DNA charge (that is, to the number of bases) and to the field strength: the longer the DNA and the stronger the field, the harder it should be to release the molecule. (This mechanism is not unlike the trapping effect on open-circular double-stranded DNA in agarose gels².)

To investigate this trapping hypothesis, we used FIGE in 6% polyacrylamide at 60 V per cm. As expected, the longer the

reverse pulse, the larger the avidinated fragments that enter the gel. In an 800/200 ms FIGE (forward/reverse pulses), all the protein-DNA complexes above 1,300 bases show the same migration velocity (Fig. 3a). Shortening the reverse pulse to 800/20 ms prevents the largest molecules from entering the gel, which now only displays sizes out to 2,686 bases (Fig. 3b). If the reverse pulse time is decreased by another order of magnitude to 2 ms, only molecules smaller than 1,500 bases can run on the gel. These FIGE regimes have little or no effect on the mobility of naked DNA. Figure 4a shows the mobility curves for a number of these experiments.

The forward pulse of 800 ms is long enough to trap modified DNA of all sizes above 1,300 bases, and the reverse pulse of 200 ms is long enough to release them. Large naked DNA shows a net motion of 3.3 μm per 800/200 ms FIGE cycle (a recoil of 1.1 μm and a forward motion of 4.4 μm). The modified molecules move a net 1.9 μm per cycle; if they have also recoiled

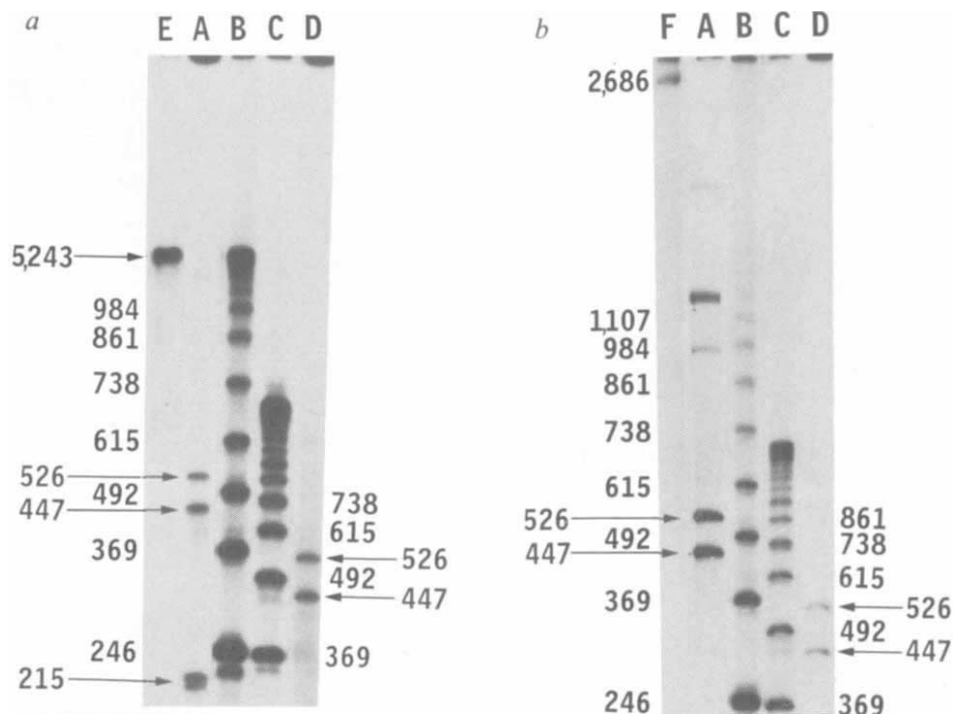


FIG. 3 Field-inversion electrophoresis of streptavidin-labelled and naked single-stranded DNA at 60 V per cm in 6% gels. Experimental procedures were as described in the legend to Fig. 1. The voltage was periodically reversed with forward/reverse pulse durations of a, 800/200 ms; and b, 800/20 ms. Lanes A, B, C and D are as in Fig. 1; lane E is SV40/*Bam*HI (5.243 kilobases); lane F is pUC19/*Hind*III (2.686 kilobases).

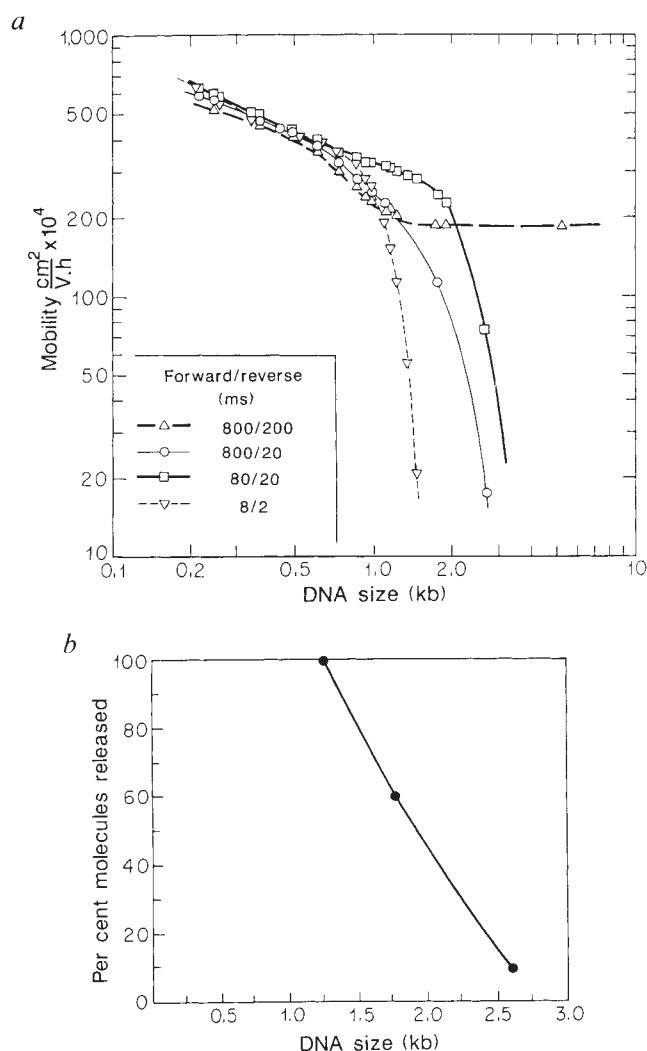


FIG. 4 *a*, The DNA size-dependence of the mobility of streptavidin-labelled single-stranded DNA in different field inversion regimes (insert) on 6% polyacrylamide gel at 60 V per cm, including those in Fig. 3. The net running time for the purpose of mobility calculation was the total forward minus total reverse pulse time. The temperature of all gels was kept constant at $\sim 15^\circ\text{C}$. *b*, The probability of the release of the streptavidin-DNA complex by 20-ms reverse pulse as a function of the DNA length. This plot is a ratio of the 800/20 ms to the 800/200 ms curves of Fig. 4*a*.

1.1 μm on release, they are trapped again after moving forward a distance, D , of 3 μm in the 6% gel. The DNA size independence of D suggests that the trapping reflects the properties of the gel matrix and of the streptavidin, rather than those of the DNA chain leading the way. The average trapping distance, D , scales as the inverse square of the acrylamide concentration, falling to 1.8 μm in an 8% gel and 0.4 μm in a 16% gel (data not shown).

The differences in mobility shown in Fig. 3*a* and *b* are due to the different reverse pulses. The 200 ms pulse releases all the molecules, but the 20-ms pulse releases only a fraction, depending on the DNA size. This fraction drops from 100% at 1,230 bases to near 0% at 2,700 bases (Fig. 4*b*).

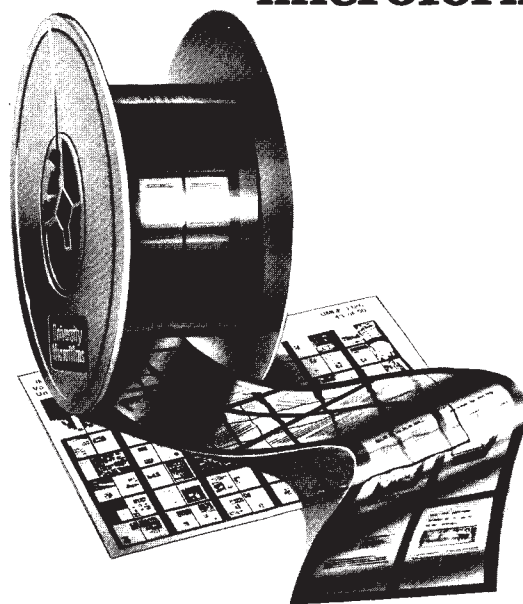
Overall, DNA trapping electrophoresis offers a controlled way to increase the size resolution of DNA molecules. ☐

Received 25 September; accepted 17 November 1989.

1. Lai, E., Davi, N. & Hood, L. *Electrophoresis* **10**, 65-67 (1989).
2. Levene, S. D. & Zimm, B. H. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4054-4057 (1987).

ACKNOWLEDGEMENTS. We thank Drs D. Sen, O. Ohara, A. Wolanski, E. Trifonov and M. Burmeister. This work was supported by a Chaim Weizmann Postdoctoral Fellowship to L.U., an NSERC Postdoctoral Fellowship to G.D., and a grant from the NIH.

**nature
is available in
microform.**



University Microfilms

International reproduces this publication in microform: microfiche and 16mm or 35mm film. For information about this publication or any of the more than 13,000 titles we offer, complete and mail the coupon to: University Microfilms International, 300 N. Zeeb Road, Ann Arbor, MI 48106. Call us toll-free for an immediate response: 800-521-3044. Or call collect in Michigan, Alaska and Hawaii: 313-761-4700.

☐ Please send information about these titles:

Name _____

Company/Institution _____

Address _____

City _____

State _____ Zip _____

Phone () _____

**University
Microfilms
International**